



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 188272**

**TO: Konstantina Katcheves**  
**Location: REM/2A60/2C70**  
**Art Unit: 1636**  
**Thursday, February 23, 2006**  
**Case Serial Number: 10/070574**

**From: Toby Port**  
**Location: Biotech-Chem Library**  
**REM-1A59**  
**Phone: 571-272-2523**

**[toby.port@uspto.gov](mailto:toby.port@uspto.gov)**

### **Search Notes**

**Examiner Katcheves,**

**See attached results.**

**If you have any questions about this search feel free to contact me at any time.**

**Thank you for using STIC search services!**

**Toby Port**  
**X22523**



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor  
Remsen Bldg. 01 D86  
571-272-2507

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☒ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library, Remsen Bldg.



STIC-Biotech/ChemLib

180572

---

**From:** Chan, Christina  
**Sent:** Wednesday, February 22, 2006 3:47 PM  
**To:** Katcheves, Konstantina; STIC-Biotech/ChemLib  
**Subject:** RE: After final 10070574

Please rush. Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
(571)-272-0841  
Remsen, 3E89

-----Original Message-----

**From:** Katcheves, Konstantina  
**Sent:** Wednesday, February 22, 2006 10:08 AM  
**To:** Chan, Christina  
**Subject:** After final 10070574

Christina,

Would you approve the following search:

Please search nucleotides 574-1134 of SEQ ID NO:1 and nucleotides 28-542<sup>1</sup> of SEQ ID NO:2 against the commercial and interference databases.

Thanks,  
Tina

***Konstantina Katcheves***  
***Patent Examiner , AU1636***  
***Phone: (571) 272-0768***  
***Room: REM 2A60***  
***Mail: REM 2C70***

2/22/2006

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:21:30 ; Search time 4624.57 Seconds  
(without alignments)  
6317.898 Million cell updates/sec

Title: US-10-070-574-2\_COPY\_28\_541

Perfect score: 514

Sequence: 1 ttattacggaagaagat.....aatgatacatgtctctgcag 514

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_sts.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_htg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 514   | 100.0       | 541    | 6     | AX099124 Sequence |
| 2          | 513   | 99.8        | 2198   | 15    | KMA376            |
| 3          | 127.8 | 24.9        | 110000 | 15    | CR382121_08       |
| 4          | 127.8 | 24.9        | 110000 | 15    | CR382121_09       |
| 5          | 73.6  | 14.3        | 2179   | 15    | AB059425          |
| 6          | 51.8  | 10.1        | 196665 | 9     | AC141560          |
| 7          | 50.8  | 9.9         | 349980 | 6     | AX344557          |
| 8          | 49.4  | 9.6         | 591    | 6     | CQ522636          |
| 9          | 49.4  | 9.6         | 3470   | 6     | CQ412028          |
| 10         | 49.4  | 9.6         | 3470   | 6     | CQ492626          |
| 11         | 49.4  | 9.6         | 3470   | 6     | CQ492817          |
| 12         | 49    | 9.5         | 217947 | 14    | AC162562          |
| 13         | 49    | 9.5         | 349980 | 6     | AX344559          |
| 14         | 48.8  | 9.5         | 16127  | 6     | AX345646          |
| 15         | 48.2  | 9.4         | 349980 | 6     | AX344566          |
| 16         | 48    | 9.3         | 349980 | 6     | AX344565          |
| 17         | 47.8  | 9.3         | 18997  | 6     | AX345472          |
| 18         | 47.8  | 9.3         | 18997  | 6     | AX347384          |

|      |      |     |        |    |            |                    |           |
|------|------|-----|--------|----|------------|--------------------|-----------|
| 19   | 47.8 | 9.3 | 18997  | 6  | AX349105   | AX349105           | Sequence  |
| 20   | 47.8 | 9.3 | 18997  | 6  | AX657772   | AX657772           | Sequence  |
| 21   | 47.8 | 9.3 | 18997  | 6  | AX659046   | AX659046           | Sequence  |
| 22   | 47.8 | 9.3 | 141466 | 9  | AC144817   | AC144817           | Mus muscu |
| 23   | 47.8 | 9.3 | 207997 | 9  | AC159136   | AC159136           | Mus muscu |
| C 24 | 47.8 | 9.3 | 349751 | 2  | PMAL4P3    | AL035476           | Plasmodiu |
| 25   | 47.6 | 9.3 | 1522   | 2  | AY701231   | AY701231           | Orconecte |
| 26   | 47.6 | 9.3 | 8392   | 6  | AX346392   | AX346392           | Sequence  |
| C 27 | 47.6 | 9.3 | 158574 | 14 | CR932984   | CR932984           | Danio rer |
| C 28 | 47.6 | 9.3 | 159017 | 5  | BX511209   | BX511209           | Zebrafish |
| C 29 | 47.4 | 9.2 | 110000 | 14 | AL954295_4 | Continuation (5 of |           |
| 30   | 47.4 | 9.2 | 250823 | 2  | AE014821   | AE014821           | Plasmodiu |
| C 31 | 47.2 | 9.2 | 1869   | 8  | BC064898   | BC064898           | Homo sapi |
| 32   | 47.2 | 9.2 | 3036   | 6  | CQ807305   | CQ807305           | Sequence  |
| 33   | 47.2 | 9.2 | 6289   | 6  | AX598860   | AX598860           | Sequence  |
| 34   | 47.2 | 9.2 | 6289   | 6  | AX599006   | AX599006           | Sequence  |
| 35   | 47.2 | 9.2 | 9289   | 6  | AX251256   | AX251256           | Sequence  |
| 36   | 47.2 | 9.2 | 9289   | 6  | AX767470   | AX767470           | Sequence  |
| 37   | 47.2 | 9.2 | 9289   | 6  | AX767546   | AX767546           | Sequence  |
| C 38 | 47.2 | 9.2 | 122746 | 15 | CR932958   | CR932958           | Medicago  |
| C 39 | 47.2 | 9.2 | 194093 | 9  | AC139037   | AC139037           | Mus muscu |
| C 40 | 47.2 | 9.2 | 198277 | 14 | CR931998   | CR931998           | Danio rer |
| 41   | 47.2 | 9.2 | 203384 | 14 | CR932357   | CR932357           | Danio rer |
| 42   | 47   | 9.1 | 5397   | 6  | AX347107   | AX347107           | Sequence  |
| C 43 | 47   | 9.1 | 215498 | 14 | AC159856   | AC159856           | Bos tauru |
| 44   | 46.8 | 9.1 | 7201   | 6  | AX345238   | AX345238           | Sequence  |
| 45   | 46.6 | 9.1 | 393    | 6  | CQ398211   | CQ398211           | Sequence  |

#### ALIGNMENTS

|            |  |   |                                   |        |     |        |                 |
|------------|--|---|-----------------------------------|--------|-----|--------|-----------------|
| RESULT 1   | AX099124   | AX099124  | Sequence 2 from Patent WO0120005. | 541 bp | DNA | linear | PAT 02-APR-2001 |
| LOCUS      | AX099124   | AX099124  | Sequence 2 from Patent WO0120005. | 541 bp | DNA | linear | PAT 02-APR-2001 |
| DEFINITION | AX099124   | AX099124  | Sequence 2 from Patent WO0120005. | 541 bp | DNA | linear | PAT 02-APR-2001 |
| ACCESSION  | AX099124   | AX099124  | Sequence 2 from Patent WO0120005. | 541 bp | DNA | linear | PAT 02-APR-2001 |
| VERSION    | AX099124.1   | GI:13538334   |                                   |        |     |        |                 |
| KEYWORDS   | Kluyveromyces marxianus                                  |   |                                   |        |     |        |                 |
| SOURCE     | Kluyveromyces marxianus                                  |   |                                   |        |     |        |                 |
| ORGANISM   | Kluyveromyces marxianus                                  |   |                                   |        |     |        |                 |
| REFERENCE  | 1  | Becher, D., Siekstele, R., Bartkeviute, D., Sasnauskas, K., Doehner, L. and Salim, S. |                                   |        |     |        |                 |
| AUTHORS    | 1  | Becher, D., Siekstele, R., Bartkeviute, D., Sasnauskas, K., Doehner, L. and Salim, S. |                                   |        |     |        |                 |
| TITLE      | Regulatory sequences and expression cassettes for yeasts |   |                                   |        |     |        |                 |
| JOURNAL    | Patent: WO 0120005-A 2 22-MAR-2001;                      |   |                                   |        |     |        |                 |
| FEATURES   | TAD Pharmazeutisches Werk GmbH (DE)                      |   |                                   |        |     |        |                 |
| source     | Location/Qualifiers                                      |   |                                   |        |     |        |                 |
|            | 1..541   |   |                                   |        |     |        |                 |
|            | /organism="Kluyveromyces marxianus"                      |   |                                   |        |     |        |                 |
|            | /mol_type="unassigned DNA"                               |   |                                   |        |     |        |                 |
|            | /db_xref="taxon:4911"                                    |   |                                   |        |     |        |                 |

#### ORIGIN

|                       |                 |  |           |             |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match           | 100.0%          | Score 514;   | DB 6;     | Length 541; |
| Best Local Similarity | 100.0%          | Pred. No. 2.6e-96;   |           |             |
| Matches 514;          | Conservative 0; | Mismatches 0;  | Indels 0; | Gaps 0;     |
| QY                    | 1               | TTTATTACGTCGAAGATAGGGAGCTTCAATGCGGTCTGTAATGGTTCATTCATT       | 60        |             |
| Db                    | 28              | TTTATTACGTCGAAGATAGGGAGCTTCAATGCGGTCTGTAATGGTTCATTCATT       | 87        |             |
| QY                    | 61              | TCGATACCTCGGGGACTTCCTTTGAATATATATCTGAGAGTATGACAGTTGGTTTCTTTC | 120       |             |
| Db                    | 88              | TCGATACCTCGGGGACTTCCTTTGAATATATATCTGAGAGTATGACAGTTGGTTTCTTTC | 147       |             |
| QY                    | 121             | TTTCTTTCTATTGTTTTTGTGTTTTATGGAATAATAGCTTTGATGATTAGGATATTTTT  | 180       |             |
| Db                    | 148             | TTTCTTTCTATTGTTTTTGTGTTTTATGGAATAATAGCTTTGATGATTAGGATATTTTT  | 207       |             |
| QY                    | 181             | GTATGTAACCAATACATGCTTGAATATATACGTAGAGGTGGGCATCTTACTCTCATTA   | 240       |             |

Db 208 FTAGTGAACAATACATGCTTGAATTAATACGTACGAGTGGGCAATCTCTCATTA 267  
Qy 241 TTGTTGTTTTTATTGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGAAACGTAA 300  
Db 268 TTGTTGTTTTTATTGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGAAACGTAA 327  
Qy 301 TATCATGTTCTCTCTTTGAAGAGGTCCCAATGCTTCCAGATAGCCAGCAATCTTT 360  
Db 328 TATCATGTTCTCTCTTTGAAGAGGTCCCAATGCTTCCAGATAGCCAGCAATCTTT 387  
Qy 361 CCATGATATTTTGGCGTTGTTTGCACGTGTCACACCTTTTCCGAACCAAGATGCAAGT 420  
Db 388 CCATGATATTTTGGCGTTGTTTGCACGTGTCACACCTTTTCCGAACCAAGATGCAAGT 447  
Qy 421 GCTGCTGATACAAACCTGTATTATCAACAATCTGGATCCATCAGCTCACAATCCACAG 480  
Db 448 GCTGCTGATACAAACCTGTATTATCAACAATCTGGATCCATCAGCTCACAATCCACAG 507  
Qy 481 CTGAAGATACAGAAAAATGATACATGCTCTCTGCAG 514  
Db 508 CTGAAGATACAGAAAAATGATACATGCTCTCTGCAG 541

RESULT 2  
KMAJ76 2198 bp DNA linear PLN 15-APR-2005  
LOCUS Kluveromyces marxianus epg1 gene encoding endopolygalacturonase,  
DEFINITION partial.  
ACCESSION AJ000076  
VERSION AJ000076.1 GI:2597956  
KEYWORDS endopolygalacturonase; epg1 gene.  
SOURCE Kluveromyces marxianus  
ORGANISM Kluveromyces marxianus  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluveromyces.  
REFERENCE 1 Siekstele,R., Bartkeviciute,D. and Sasnauskas,K.  
AUTHORS Cloning, targeted disruption and heterologous expression of the  
TITLE Kluveromyces marxianus endopolygalacturonase gene (EPG1)  
JOURNAL Yeast 15 (4), 311-322 (1999)  
PUBMED 10206190  
REFERENCE 2 (bases 1 to 2198)  
AUTHORS Siekstele,R.  
TITLE Direct Submission  
JOURNAL Submitted (07-JUL-1997) Siekstele R., Gene engineering laboratory,  
Institute of Biotechnology, V.Graiciuno 8, 2028, LITHUANIAN  
REPUBLIC

REMARK revised by author 04-NOV-97  
FEATURES  
Location/Qualifiers  
1..2198  
/organism="Kluveromyces marxianus"  
/mol\_type="genomic DNA"  
/strain="BMK Y-719"  
/db\_xref="taxon:4911"  
/chromosome="VI"  
338..>1655  
/gene="epg1"  
338..352  
/gene="epg1"  
573..>1655  
/gene="epg1"  
/EC\_number="3.2.1.15"  
/codon\_start=1  
/product="endopolygalacturonase"  
/protein\_id="CAA03900.1"  
/db\_xref="GOA:O13478"  
/db\_xref="InterPro:IPR000743"  
/db\_xref="InterPro:IPR006626"  
/db\_xref="UniProt/TREMBL:O13478"  
/translation="MLFSNTLLIAASALLAASPLEKRDSTLSGKTAGGSLNCAT  
VTNNVVEPSGKTLDTLPGDGTAFVGVTFGYDEWVGPLVLSIGRNKIKVVGKSGH

LLDGDGARWWDGKGKKPKFMSLRLLTGNSDVGGLOIKNTPIQALISVNSCSDTVIH  
DVTIDNRDDKDLNHLGHTYDGFYGVNNTVITENSHVYNQDDCLAVNSGTGVYKNNYC  
SGHGASIGSVGLRNNVVDIVIFENNOIVNSNGLAIKTIQAKATSGVNNVHFLSNTI  
SGIRKFGIVVETDYSNGSTTGTGSKVPITNFEDVGLTGSVDSEAYRKVIFVAGASKW  
TKNVDITGSSSFSGCTGIPSGGAFG"  
573..657  
/gene="epg1"  
658..>1655  
/gene="epg1"  
/product="endopolygalacturonase"  
/EC\_number="3.2.1.15"  
sig\_peptide  
mat\_peptide  
ORIGIN  
Query Match 99.8%; Score 513; DB 15; Length 2198;  
Best Local Similarity 100.0%; Pred. No. 2.9e-96;  
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TTTATTAAACGTGAAGAAGATAAGGGAAGTCTTCAATCGGTTCTGAATGGTTGATCCATT 60  
Db 1686 TTTATTAAACGTGAAGAAGATAAGGGAAGTCTTCAATCGGTTCTGAATGGTTGATCCATT 1745  
Qy 61 TCGATACCTCGGGGACTTCTTGAATATATCTCGAGAGTATGACAGTTGGTTCTTCTTC 120  
Db 1746 TCGATACCTCGGGGACTTCTTGAATATATCTCGAGAGTATGACAGTTGGTTCTTCTTC 1805  
Qy 121 TTTCTTTCTATTGTTTGTGTTTATGAAATATAGCTTTTCATGATTTAGGATATTTTTT 180  
Db 1806 TTTCTTTCTATTGTTTGTGTTTATGAAATATAGCTTTTCATGATTTAGGATATTTTTT 1865  
Qy 181 GTAGTGAACCAATACATGCTTTGATTATATACGTACGAGTGGGCAATCTTACTCTCATTA 240  
Db 1866 GTAGTGAACCAATACATGCTTTGATTATATACGTACGAGTGGGCAATCTTACTCTCATTA 1925  
Qy 241 TTGTTGTTTTTATGAGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGACGTAA 300  
Db 1926 TTGTTGTTTTTATGAGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGACGTAA 1985  
Qy 301 TATCCATGTTCTTCTCTTTGAAGAGTCCCAACCAATGCTTCCAGATAGCCAGCAATCTT 360  
Db 1986 TATCCATGTTCTTCTCTTTGAAGAGTCCCAACCAATGCTTCCAGATAGCCAGCAATCTT 2045  
Qy 361 CCATGATATTTTGGCGTTGTTTGCACGTGTCACACCTTTTCGAACCAAGATGTCAGT 420  
Db 2046 CCATGATATTTTGGCGTTGTTTGCACGTGTCACACCTTTTCGAACCAAGATGTCAGT 2105  
Qy 421 GCTGCTGATACAAACCTGTATTATCAACAATCTGGATCCATCAGCTCACAATCCACAG 480  
Db 2106 GCTGCTGATACAAACCTGTATTATCAACAATCTGGATCCATCAGCTCACAATCCACAG 2165  
Qy 481 CTGAAGATACAGAAAAATGATACATGCTCTCTGCA 513  
Db 2166 CTGAAGATACAGAAAAATGATACATGCTCTCTGCA 2198

RESULT 3  
CR382121.08/c  
WFOCOMMENT  
Sequence split into 11 fragments LOCUS CR382121 Accession CR382121  
Fragment Name Begin End  
CR382121\_00 1 110000  
CR382121\_01 100001 210000  
CR382121\_02 200001 310000  
CR382121\_03 300001 410000  
CR382121\_04 400001 510000  
CR382121\_05 500001 610000  
CR382121\_06 600001 710000  
CR382121\_07 700001 810000  
CR382121\_08 800001 910000  
CR382121\_09 900001 1010000  
CR382121\_10 1000001 1062590  
Continuation (9 of 11) of CR382121 from base 800001 (CR382121 Kluveromyces lactis str:  
Query Match 24.9%; Score 127.8; DB 15; Length 110000;



```

REFERENCE
AUTHORS      Parkway, St. Louis, MO 63108, USA
TITLE        3 (bases 1 to 196665)
JOURNAL      Wilson.R.K.
SUBMITTED    Submitted (16-JUN-2004) Genome Sequencing Center, 4444 Forest Park
JOURNAL      Parkway, St. Louis, MO 63108, USA
REFERENCE    4 (bases 1 to 196665)
AUTHORS      Wilson.R.K.
TITLE        Direct Submission
JOURNAL      Submitted (02-SEP-2004) Genome Sequencing Center, 4444 Forest Park
SUBMITTED    Parkway, St. Louis, MO 63108, USA
REFERENCE    5 (bases 1 to 196665)
AUTHORS      Wilson.R.K.
TITLE        Direct Submission
JOURNAL      Submitted (27-JAN-2005) Genome Sequencing Center, Washington
SUBMITTED    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL      MO 63108, USA
COMMENT      On Sep 2, 2004 this sequence version replaced gi:48762631.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: M_BA0425A05
-----

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC library has been constructed by Kazutoyo Osegawa
and Minako Tatenno in the laboratory of Pieter de Jong
(http://www.chori.org) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (http://www.resgen.com) or
Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
FEATURES
source
    Location/Qualifiers
        1..196665
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="12"
            /map="12"
            /clone="RP23-425A5"
            /clone_lib="RPCI-23"
repeat_region
    1..55
        /rpt_family="Alu"
repeat_region
    68..248
        /rpt_family="B2"
repeat_region
    299..668
        /rpt_family="L1"
repeat_region
    674..879
        /rpt_family="L1"
repeat_region
    994..1094
        /rpt_family="B2"
repeat_region
    1116..1234
        /rpt_family="L1"
repeat_region
    1574..1642
        /rpt_family="B4"
repeat_region
    2798..2936
        /rpt_family="B2"
repeat_region
    3148..3215
        /rpt_family="Alu"
repeat_region
    3216..3229
        /rpt_family="ID"
repeat_region
    3825..3955
        /rpt_family="B4"
repeat_region
    4010..4155
        /rpt_family="MER1_type"
repeat_region
    4275..4422
        /rpt_family="Alu"
repeat_region
    5113..5173
        /rpt_family="Alu"
repeat_region
    5653..5768
        /rpt_family="MER1_type"
repeat_region
    5999..6141
        /rpt_family="Alu"
repeat_region
    6151..6343
        /rpt_family="B2"
repeat_region
    6494..6618
        /rpt_family="Alu"
repeat_region
    6619..6754
        /rpt_family="B4"
repeat_region
    6786..6800
        /rpt_family="B4"
repeat_region
    6801..6864
        /rpt_family="ID"
repeat_region
    6993..7182
        /rpt_family="Alu"
repeat_region
    7627..7849
        /rpt_family="MER1_type"
repeat_region
    8112..8201
        /rpt_family="ERV1"
repeat_region
    8344..8699
        /rpt_family="MaLR"
repeat_region
    8952..9007
        /rpt_family="ERV1"
repeat_region
    9724..9760
        /rpt_family="MaLR"
repeat_region
    9761..9881
        /rpt_family="ERV1"
repeat_region
    9893..10224
        /rpt_family="MaLR"
repeat_region
    10556..10694
        /rpt_family="B2"
repeat_region
    11080..11282
        /rpt_family="B2"
repeat_region
    11324..11495
        /rpt_family="B2"
repeat_region
    11518..11616
        /rpt_family="B4"
repeat_region
    11985..12519
        /rpt_family="MaLR"
repeat_region
    12768..12832
        /rpt_family="L1"
repeat_region
    13018..13137
        /rpt_family="L1"
repeat_region
    13163..13218
        /rpt_family="Alu"
repeat_region
    16855..17063
        /rpt_family="B2"
repeat_region
    17192..17455
        /rpt_family="B4"
repeat_region
    18530..18636
        /rpt_family="Alu"
repeat_region
    18732..18803
        /rpt_family="B4"
repeat_region
    19179..19988
        /rpt_family="L1"

```







Db 3129 TTTTTCGTTTATTAATACGACGAGTGG 223  
Qy 197 TCGTTGTTTATTAATACGACGAGTGG 223  
Db 3069 TCGTTGTTTATTAATACGACGAGTGG 3043

RESULT 12  
AC162562/c  
LOCUS AC162562.2 GI:69301550  
DEFINITION HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
ACCESSION AC162562.2  
VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus

AC162562 217947 bp DNA linear HTG 01-JUL-2005  
Bos taurus clone CH240-120C4, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 17  
unrounded pieces.  
AC162562  
AC162562.2 GI:69301550  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Bos taurus (cow)  
Bos taurus

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 217947)  
Muzny D.Marie, Metzker M.Lee, Abramson S., Adams C., Alder J.,  
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,  
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,  
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,  
Biswal K., Blair J., Blankenbush K., Blyth P., Brown M.,  
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,  
Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,  
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,  
Cleveland C., Cockrell R., Cox C., Coylie W., Cree A., D'Souza D.,  
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,  
Delgado O., Denison S., Deramo C., Ding Y., Dinh H., Divya K.,  
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,  
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,  
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,  
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,  
Gubregoriep E., Geer K., Gill R., Grady K., Guerra W., Guevara W.,  
Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K.,  
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,  
Hernandez R., Hines S., Hladun S., Hodgson A., Hognes M.,  
Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,  
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,  
Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,  
Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,  
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,  
Lorenshewa L., Loulsged H., Lozado R.J., Lu X., Ma J.,  
Maheshwari M., Mahandartine M., Mahmoud M., Malloy K., Mangum A.,  
Mangum B., Mapus P., Martin K., Martin R., Martinez E.,  
Mawhinney S., McLeod M.P., McNeill T.Z., Meenen E.,  
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,  
Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,  
Nankervis C., Neal D., Newton N., Nguyen N., Norris S., Parks K.,  
Nwaokemele O., Okwunonu G., Olarnpusoon A., Pal S., Parks K.,  
Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,  
Plapper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,  
Puzo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,  
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,  
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,  
Sanders W., Savery G., Scherer S., Scott G., Shatsman S., Shen H.,  
Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smales D.,  
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,  
Steimle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C.,  
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Umani K.,  
Valas R., Vera V., Villaana D., Waldron L., Walker B., Wang J.,  
Wang C., Wang S., Warren J., Warren R., Wei X., White F.,  
Williams G., Willson R., Wleczky R., Wooden H., Worley K.,  
Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,  
Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von  
Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,  
Weinstock G. and Gibbs R.A.  
Direct Submission  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 217947)  
Worley K.C.  
Direct Submission  
Submitted (30-MAY-2005) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 217947)  
Cow Genome Sequencing Consortium.  
Direct Submission  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 29, 2005 this sequence version replaced gi:66793710.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly ('contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: FHI0  
Center clone name: CH240-120C4  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 211614 bases at least Q40  
Consensus quality: 213060 bases at least Q30  
Consensus quality: 214433 bases at least Q20  
Estimated insert size: 217550; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 14309: contig of 14309 bp in length  
\* 14310 14359: gap of 50 bp  
\* 14360 45303: contig of 30944 bp in length  
\* 45304 45376: gap of 73 bp  
\* 45377 53807: contig of 8431 bp in length  
\* 53808 54000: gap of 193 bp  
\* 54001 66596: contig of 12596 bp in length  
\* 66597 67194: gap of 598 bp  
\* 67195 90532: contig of 23338 bp in length  
\* 90533 90582: gap of 50 bp  
\* 90583 111478: contig of 20896 bp in length  
\* 111479 111528: gap of 50 bp  
\* 111529 113900: contig of 2372 bp in length  
\* 113901 113950: gap of 50 bp  
\* 113951 122318: contig of 8368 bp in length  
\* 122319 122431: gap of 113 bp  
\* 122432 149282: contig of 26851 bp in length  
\* 149283 149696: gap of 414 bp  
\* 149697 160191: contig of 10495 bp in length  
\* 160192 160258: gap of 66 bp  
\* 160259 177355: contig of 17098 bp in length  
\* 177356 177405: gap of 50 bp

TITLE  
JOURNAL





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:20:52 ; Search time 450.886 Seconds  
(without alignments)  
7597.608 Million cell updates/sec

Title: US-10-070-574-2\_COPY\_28\_541

Perfect score: 514

Sequence: 1 ttattacgtgaagaagat.....aatgatactgtcttcgag 514

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_21.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 514   | 100.0       | 541    | 5     | Aaf61990 K. marxia |
| C 2        | 49.4  | 9.6         | 591    | 5     | ABV54484 Human pro |
| C 3        | 49.4  | 9.6         | 3470   | 5     | ABV24695 Human pro |
| C 4        | 49.4  | 9.6         | 3470   | 5     | ABV24504 Human pro |
| C 5        | 49.4  | 9.6         | 3470   | 5     | ADL45209 Human ova |
| 6          | 48.8  | 9.5         | 16127  | 6     | ABL32744 Human imm |
| 7          | 47.8  | 9.3         | 18997  | 6     | ABL32570 Human imm |
| 8          | 47.8  | 9.3         | 18997  | 6     | ABK33948 Human DNA |
| 9          | 47.8  | 9.3         | 18997  | 8     | ADA20352 Prostate  |
| 10         | 47.8  | 9.3         | 18997  | 8     | ADA84159 Human ren |
| 11         | 47.6  | 9.3         | 8392   | 6     | ABL33490 Human imm |
| C 12       | 47.2  | 9.2         | 2000   | 11    | ACL37108 Rice stre |
| 13         | 47.2  | 9.2         | 3036   | 13    | ADS89739 Oligonuc  |
| 14         | 47.2  | 9.2         | 6289   | 8     | ABZ10206 Haematopo |
| 15         | 47.2  | 9.2         | 6289   | 8     | ABZ10060 Haematopo |
| 16         | 47.2  | 9.2         | 9289   | 4     | AAS46502 Tumour su |
| 17         | 47.2  | 9.2         | 9289   | 10    | ADE84198 Human lym |
| 18         | 47.2  | 9.2         | 9289   | 10    | ADE84122 Human lym |
| 19         | 47    | 9.1         | 5397   | 6     | ABL34205 Human imm |

|      |      |     |        |    |                    |
|------|------|-----|--------|----|--------------------|
| 20   | 45.8 | 9.1 | 7201   | 6  | ABL32336 Human imm |
| 21   | 45.6 | 9.1 | 393    | 5  | ADI72540 Human ova |
| 22   | 46.6 | 9.1 | 393    | 5  | ADI37679 Human ova |
| 23   | 46.6 | 9.1 | 497    | 5  | ADI37664 Human ova |
| 24   | 46.6 | 9.1 | 497    | 5  | ADI72525 Human ova |
| 25   | 46.6 | 9.1 | 5474   | 6  | ABL33270 Human imm |
| 26   | 46.6 | 9.1 | 8147   | 6  | ABL32428 Human imm |
| 27   | 46.6 | 9.1 | 113515 | 6  | ABL34174 Human imm |
| 28   | 46.2 | 9.0 | 6316   | 6  | ABL33639 Human imm |
| 29   | 46   | 8.9 | 6182   | 6  | ABL49388 Human pol |
| C 30 | 45.6 | 8.9 | 15649  | 6  | ABL70544 Chemical  |
| C 31 | 45.6 | 8.9 | 5949   | 14 | ADY18488 DNA encod |
| C 32 | 45.4 | 8.8 | 2000   | 11 | ACL35887 Rice stre |
| 33   | 45.4 | 8.8 | 5771   | 6  | ABN80073 Human che |
| 34   | 45.4 | 8.8 | 9905   | 6  | ABL32062 Human imm |
| C 35 | 45.2 | 8.8 | 2355   | 12 | ADP18734 Mouse lip |
| 36   | 45.2 | 8.8 | 5034   | 13 | ADS89648 Oligonuc  |
| 37   | 45.2 | 8.8 | 6577   | 4  | AAS46718 Tumour su |
| 38   | 45.2 | 8.8 | 10286  | 4  | AAS45308 Chemical  |
| C 39 | 45.2 | 8.8 | 10286  | 6  | ABK28147 DNA trans |
| C 40 | 45   | 8.8 | 526    | 6  | ABQ39216 Oligonuc  |
| 41   | 45   | 8.8 | 526    | 6  | ABQ39216 Oligonuc  |
| C 42 | 45   | 8.8 | 1221   | 11 | ADL33377 Human tra |
| 43   | 45   | 8.8 | 3326   | 13 | ADS89718 Oligonuc  |
| 44   | 45   | 8.8 | 3326   | 13 | ADS89444 Oligonuc  |
| 45   | 45   | 8.8 | 7316   | 6  | ABN80240 Human che |

## ALIGNMENTS

## RESULT 1

AAF61990 ID AAF61990 standard; DNA; 541 BP.

XX AAF61990;

DT 22-AUG-2001 (first entry)

XX K. marxianus endopolygalacturonase terminator region DNA SEQ ID 2.  
XX Endopolygalacturonase; promoter; hepatitis B surface antigen; pectin;  
XX polyoma virus VP1; staphylococcal protein A; vaccine; antiviral;  
XX antibacterial; regulatory region; terminator; ds.

OS Kluyveromyces marxianus.

XX WO200120005-A1.

XX 22-MAR-2001.

XX 05-SEP-2000; 2000WO-EP008662.

XX 10-SEP-1999; 99DE-01043383.

XX (TADP-) TAD PHARMA GMBH.

XX Becher D, Siekstele R, Bartkeviciute D, Sasnauskas K, Doehner L;  
XX Salim S;  
XX WPI; 2001-244812/25.

XX New promoter from Kluyveromyces marxianus, useful for controlling  
XX expression of foreign genes in yeast, e.g. production of vaccine  
XX antigens.

XX Claim 2; Page 27; 32pp; German.

XX This invention describes a novel promoter sequence (S1), isolated from  
XX Kluyveromyces marxianus. (S1) is a promoter for controlling expression of  
XX foreign proteins in yeast, particularly hepatitis B surface antigen,  
XX polyoma virus VP1 or staphylococcal protein A for use in vaccines. (S1)  
XX is a strong promoter that can be induced by adding pectin to the culture

CC medium and when combined with its native signal sequence can provide  
 CC secretion of heterologous protein. Cassettes based on (S1) are stably  
 CC retained in the chromosome when cells are grown under optimal conditions  
 CC and provide high yields of recombinant protein. Kluyveromyces marxianus  
 CC is generally regarded as safe, can use a wide variety of carbon sources,  
 CC is not very sensitive to temperature (it can grow at up to 45 plusOC) and  
 CC under ideal conditions has a doubling time of only 35 minutes. The  
 CC products of the invention have antiviral and antibacterial activity. This  
 CC sequence represents a terminator region of the Kluyveromyces marxianus  
 CC endopolygalacturonase described in the disclosure of the invention  
 XX  
 SQ Sequence 541 BP; 138 A; 96 C; 104 G; 203 T; 0 U; 0 Other;

Query Match 100.0%; Score 514; DB 5; Length 541;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-114;  
 Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 TTTATTAACTGAAGAAGATAAGGAAGTCTTCAATCGGTTCTGAAATGGTTGATCCATT 60  
 Db TTTATTAACTGAAGAAGATAAGGAAGTCTTCAATCGGTTCTGAAATGGTTGATCCATT 87  
 Qy 61 TCGATACCTCGGGACTTCTTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTTC 120  
 Db TCGATACCTCGGGACTTCTTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTTC 147  
 Qy 121 TTTCTTTCTATTGTTTTTTTATGGAATATAGCTTTGATCATTTAGGATATTTTTT 180  
 Db TTTCTTTCTATTGTTTTTTTATGGAATATAGCTTTGATCATTTAGGATATTTTTT 207  
 Qy 181 GTAGTGAACCAATACATGCTTTGATTAATATACGTACGAGTGGGCACTTCTACTCTCATTA 240  
 Db GTAGTGAACCAATACATGCTTTGATTAATATACGTACGAGTGGGCACTTCTACTCTCATTA 267  
 Qy 241 TTGGTGTTTTATTGGAGGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGCGAAACGTAA 300  
 Db TTGGTGTTTTATTGGAGGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGCGAAACGTAA 327  
 Qy 301 TATCCATGTTCTTCTTTCGAAGAGTCCACCAATGCTTCCAGATAGCCAGCATTCCT 360  
 Db TATCCATGTTCTTCTTTCGAAGAGTCCACCAATGCTTCCAGATAGCCAGCATTCCT 387  
 Qy 361 CCATGATATTTTGGCTTGTGTTTGGCACTGGTGACACCCCTTCGAACCAAAAGATGTCAAGT 420  
 Db CCATGATATTTTGGCTTGTGTTTGGCACTGGTGACACCCCTTCGAACCAAAAGATGTCAAGT 447  
 Qy 421 GCTGCTGATACAAACCTGTATTCATACAAATTCGTGATCATCAGCTCAATCCACAG 480  
 Db GCTGCTGATACAAACCTGTATTCATACAAATTCGTGATCATCAGCTCAATCCACAG 507  
 Qy 481 CTGAAGATACAGAAATGATACATGCTCTGCGAG 514  
 Db CTGAAGATACAGAAATGATACATGCTCTGCGAG 541

RESULT 2  
 ID ABV54484/c  
 XX ABV54484 standard; cDNA; 591 BP.  
 AC  
 AC ABV54484;  
 XX  
 DT 17-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 54475.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.

PF 20-FEB-2001; 2001WO-US005171.  
 XX  
 PR 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PA  
 XX Schlegel R, Endege WO, Monahan JB;  
 PI  
 XX WPI; 2001-662795/76.  
 XX  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer.  
 XX  
 PS Claim 1; Page 10531; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for: (a) assessing whether  
 CC a patient is afflicted with prostate cancer; (b) monitoring the  
 CC progression of prostate cancer in a patient; (c) assessing the efficacy  
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
 CC determining whether prostate cancer has metastasized in a patient; (h)  
 CC assessing the aggressiveness or indolence of prostate cancer in a patient  
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
 CC  
 SQ Sequence 591 BP; 230 A; 129 C; 120 G; 112 T; 0 U; 0 Other;  
 XX  
 Query Match 9.6%; Score 49.4; DB 5; Length 591;  
 Best Local Similarity 58.5%; Pred. No. 0.041;  
 Matches 86; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
 Qy 77 TTCCTTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTCTTCTTCTTCTTCTTCTT 136  
 Db 303 TTTTCTTAATT 244  
 Qy 137 TTGTTTTTATGGAATATATAGCTTTGATGATTTAGGATATTTTTTTAGTGAAACCAATACA 196  
 Db 243 TTTTCTT 184  
 Qy 197 TGCTTGATTAATATATACGTACGAGGTGG 223  
 Db 183 TGCTGATTTATTACAGGGATAAGATGG 157  
 XX  
 RESULT 3  
 ABV24695/c  
 ID ABV24695 standard; cDNA; 3470 BP.  
 XX  
 AC ABV24695;  
 XX  
 DT 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 24686.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.



















GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:40:58 ; Search time 3102.17 Seconds  
(without alignments)  
7752.175 Million cell updates/sec

Title: US-10-070-574-2\_COPY\_28\_541

Perfect score: 514

Sequence: 1 ttattacacgtgaagaat.....aatgatacatgtctctcag 514

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_est3.\*

4: gb\_est4.\*

5: gb\_est5.\*

6: gb\_est6.\*

7: gb\_est7.\*

8: gb\_est8.\*

9: gb\_gss1.\*

10: gb\_gss2.\*

11: gb\_gss3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID       | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 87.2  | 17.0        | 515    | 9 AZ928077  | AZ928077 479.dif04 |
| 2          | 77.8  | 15.1        | 986    | 11 CNS06N3D | AL406271 T3 end of |
| 3          | 53.8  | 10.5        | 779    | 11 CNS04A9H | AL281582 Tetraodon |
| 4          | 53.6  | 10.4        | 950    | 11 CNS07185 | AL424587 T3 end of |
| 5          | 52    | 10.1        | 975    | 9 BH179465  | BH179465 014_P_10- |
| 6          | 49.8  | 9.7         | 767    | 10 CNS00A0X | AL055924 Drosophil |
| 7          | 49.6  | 9.6         | 1051   | 8 DN569917  | DN569917 94034428  |
| 8          | 49.4  | 9.6         | 1175   | 2 BE421167  | BE421167 HWM006.E0 |
| 9          | 48.8  | 9.5         | 325    | 10 CNS0099P | AL053301 Drosophil |
| 10         | 48.8  | 9.5         | 1094   | 10 CNS012FZ | AL101513 Drosophil |
| 11         | 48.8  | 9.5         | 1101   | 10 CNS001GJ | AL050877 Drosophil |
| 12         | 48.4  | 9.4         | 452    | 2 B1142777  | B1142777 rK73g10.Y |
| 13         | 48.4  | 9.4         | 1083   | 9 BZ696101  | BZ696101 SP_Ba007  |
| 14         | 48.4  | 9.4         | 1201   | 10 CNS0167M | AL106396 Drosophil |
| 15         | 48    | 9.3         | 475    | 9 BH843983  | BH843983 TC3-55H10 |
| 16         | 48    | 9.3         | 673    | 9 BH842470  | BH842470 TC3-55H9  |
| 17         | 47.8  | 9.3         | 759    | 10 CNS007BZ | AL066974 Drosophil |
| 18         | 47.6  | 9.3         | 284    | 3 BM882253  | BM882253 rdb2e12.Y |
| 19         | 47.6  | 9.3         | 348    | 1 A1155694  | A1155694 ue03a05.r |
| 20         | 47.6  | 9.3         | 652    | 7 CO070912  | CO070912 GR_Ea28D  |
| 21         | 47.6  | 9.3         | 822    | 6 CB849204  | CB849204 MSA-0892  |
| 22         | 47.6  | 9.3         | 869    | 10 CNS017Z2 | AL108680 Drosophil |

|      |      |     |      |             |                    |
|------|------|-----|------|-------------|--------------------|
| C 23 | 47.6 | 9.3 | 919  | 10 CNS005RL | AL061409 Drosophil |
| C 24 | 47.6 | 9.3 | 1334 | 5 BU849682  | BU849682 AGENCOURT |
| C 25 | 47.4 | 9.2 | 376  | 3 BM037194  | BM037194 fu84c02.Y |
| C 26 | 47.4 | 9.2 | 724  | 9 BZ086084  | BZ086084 l1e26a01. |
| C 27 | 47.4 | 9.2 | 801  | 10 AG034099 | AG034099 Pan tregl |
| C 28 | 47.4 | 9.2 | 946  | 3 BI911499  | BI911499 603064008 |
| C 29 | 47.4 | 9.2 | 1101 | 10 CNS017SO | AL108450 Drosophil |
| C 30 | 47.2 | 9.2 | 950  | 8 DN584697  | DN584697 92277725  |
| C 31 | 47.2 | 9.2 | 1375 | 10 AG350207 | AG350207 Mus muscu |
| C 32 | 47   | 9.1 | 680  | 9 BH179930  | BH179930 015_L_18- |
| C 33 | 47   | 9.1 | 680  | 11 CNS07LWI | AL161684 T3 end of |
| C 34 | 46.8 | 9.1 | 363  | 5 BU761068  | BU761068 sage2e07. |
| C 35 | 46.6 | 9.1 | 295  | 5 BU529209  | BU529209 AGENCOURT |
| C 36 | 46.6 | 9.1 | 929  | 8 DN584074  | DN584074 93854031  |
| C 37 | 46.6 | 9.1 | 994  | 10 CNS0145S | AL103738 Drosophil |
| C 38 | 46.6 | 9.1 | 1101 | 10 CNS0039G | AL063921 Drosophil |
| C 39 | 46.6 | 9.1 | 1127 | 11 CNS06FTQ | AL396456 T7 end of |
| C 40 | 46.6 | 9.1 | 1200 | 10 CL031866 | CL031866 CH216-33I |
| C 41 | 46.4 | 9.0 | 786  | 10 CNS006O9 | AL065615 Drosophil |
| C 42 | 46.4 | 9.0 | 902  | 10 CNS010WI | AL099516 Drosophil |
| C 43 | 46.4 | 9.0 | 1089 | 10 CL511283 | CL511283 SAIL_844  |
| C 44 | 46.4 | 9.0 | 1101 | 10 CNS002C3 | AL097485 Drosophil |
| C 45 | 46.2 | 9.0 | 1101 | 10 CNS010D3 | AL098817 Drosophil |

#### ALIGNMENTS

RESULT 1  
AZ928077 515 bp DNA linear GSS 01-APR-2001  
LOCUS 479.dif04c06.s1 Saccharomyces kluyveri genomic clone 479.dif04c06.s1, genomic survey sequence.  
DEFINITION AZ928077.1 GI:13498982  
ACCESSION  
VERSION  
KEYWORDS GSS.  
SOURCE Saccharomyces kluyveri  
ORGANISM Saccharomyces kluyveri

REFERENCE 1 (bases 1 to 515)  
AUTHORS Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish,W.R., Waterston,R.H. and Johnston,M.  
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis  
JOURNAL Unpublished (2001)  
COMMENT Contact: Johnston M  
Department of Genetics  
Washington University Medical School  
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA  
Tel: 314 362 2735  
Fax: 314 362 7855  
Email: mj@genetics.wustl.edu  
Class: random plasmid subclone.

FEATURES  
source  
Location/Qualifiers  
1..515  
/organism="Saccharomyces kluyveri"  
/mol\_type="genomic DNA"  
/strain="NRRL Y-12651 (CBS 3082)"  
/db\_xref="taxon:4934"  
/clone="479.dif04c06.s1"  
/clone\_lib="Saccharomyces kluyveri"  
/note="Random genomic sequence"

#### ORIGIN

Query Match 17.0%; Score 87.2; DB 9; Length 515;  
Best Local Similarity 63.5%; Pred. No. 2 5e-10;  
Matches 148; Conservative 0; Mismatches 84; Indels 1; Gaps 1;  
QY 282 GTTTAGAGCGCAACGTAAATCCATGTTCTTCTTTGAAGAGGTCACCATTTGCTTC 341  
|||||  
DB 281 GTTTAAACTGCAACAGTTATGTCAGGTTTTTCTCCTCCAAATGTCCTCCACCAAGCTTT 340  
|||||











|                       |  |      |   |
|-----------------------|--|------|---|
| ORIGIN                |  | 201  | TGATTATATACGTACGAGTGGGCATCTCTCTCATTTATTTGGTGTGTTTATTGGGCA 260       |
| Query Match           |  |      | 9.5%; Score 48.8; DB 10; Length 325;                                |
| Best Local Similarity |  |      | 42.4%; Pred. No. 0.42;  |
| Matches               |  |      | 101; Conservative 25; Mismatches 112; Indels 0; Gaps 0;             |
| QY                    |  | 77   | TTCTTTTGAATATCTGAGAGTATGACAGTTGGTGTGTTTCTTTCTTTCTTTCTTTATGTTT 136   |
| Db                    |  | 253  | TTT 194     |
| QY                    |  | 137  | TTGTTTTTGAATATATAGCTTTGAGATTTAGGATATTTTGTAGTGAACCAATACA 196         |
| Db                    |  | 193  | TTTTTTTWTWTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 134     |
| QY                    |  | 197  | TGCTTCATTAATATACGTACGAGTGGCATCTCTCTCATTTATTTGGTGTGTTTATTGGA 256     |
| Db                    |  | 133  | WMTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 74     |
| QY                    |  | 257  | GGGAAAAATTAATCTAGGAGTATCGTTTGTAGCGGCAAGTAATATCCATGTTCTTC 314        |
| Db                    |  | 73   | AGTTTATATTAATATACAGTTGTTTGTAGAAATGTAATAATATGTTTATTTTC 16            |
| RESULT 10             |  |      |   |
| CNS012FZ/c            |  |      |   |
| LOCUS                 |  |      |   |
| DEFINITION            |  |      |   |
| ACCESSION             |  |      |   |
| VERSION               |  |      |   |
| KEYWORDS              |  |      |   |
| SOURCE                |  |      |   |
| ORGANISM              |  |      |   |
| REFERENCE             |  |      |   |
| AUTHORS               |  |      |   |
| JOURNAL               |  |      |   |
| COMMENT               |  |      |   |
| FEATURES              |  |      |   |
| source                |  |      |   |
| ORIGIN                |  |      |   |
| Query Match           |  |      | 9.5%; Score 48.8; DB 10; Length 1094;                               |
| Best Local Similarity |  |      | 37.2%; Pred. No. 0.48;  |
| Matches               |  |      | 90; Conservative 43; Mismatches 109; Indels 0; Gaps 0;              |
| QY                    |  | 81   | TTTGAATATATCTGAGAGTATGACAGTTGGTGTGTTTCTTTCTTTCTTTCTTTCTTTTGT 140    |
| Db                    |  | 1067 | TTTTTTTDDWWTTTGAAARAGAAATTTTWTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1008   |
| QY                    |  | 141  | TTTTTATGAAATATAGCTTTGATGATTTAGGATATTTTGTAGTGAACCAATACATGCT 200      |
| Db                    |  | 1007 | TTTTTTADMRRAATTTTWDRTAATAWATTTTWTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 948 |
| ORIGIN                |  |      |   |
| Query Match           |  |      | 9.5%; Score 48.8; DB 10; Length 1101;                               |
| Best Local Similarity |  |      | 37.1%; Pred. No. 0.48;  |
| Matches               |  |      | 92; Conservative 42; Mismatches 114; Indels 0; Gaps 0;              |
| QY                    |  | 108  | TTGGTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 167    |
| Db                    |  | 384  | TTTCTGTTTWTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 325     |
| QY                    |  | 168  | TAGGATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 227     |
| Db                    |  | 324  | TTTCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 265     |
| QY                    |  | 228  | TCTACTCTCATTTATTTGTTTATTTAGGCGGAAAAATTAATATAGGAGTATCGTTTATG 287     |
| Db                    |  |      |   |







```
Db 754 ATTGTTCTTAGTTTCACTACTATTATTATTATTCATATTATGATGTTATTGACATCGCGAAGCA 813
Qy 241 CCAGCCTCCATACCGATTAGACAGGATCTCAACGCTGGGCTCCAGAGCTACACATTATG 300
Db 814 CCAGCCTCCATACCGATTAGACAGGATCTCAACGCTGGGCTCCAGAGCTACACATTATG 873
Qy 301 CTAATAACTATCTACTGTAACTACAGCTACAGAAAAAACTATAAAAGAGCGAGGATAAA 360
Db 874 CTAATAACTATCTACTGTAACTACAGCTACAGAAAAAACTATAAAAGAGCGAGGATAAA 933
Qy 361 CCACCTCTCTTGTGAATCAGGATCAGTAGGTAACCTATCAATAACCTCTCTCTCTCAAA 420
Db 934 CCACCTCTCTTGTGAATCAGGATCAGTAGGTAACCTATCAATAACCTCTCTCTCTCAAA 993
Qy 421 ATATCAATAACAGTAGTATCAACAGCATATCAATAACTACTACTACACAGTAG 480
Db 994 ATATCAATAACAGTAGTATCAACAGCATATCAATAACTACTACTACACAGTAG 1053
Qy 481 GAACAGTAACGACAAACGACAGATAGTAACGACAATAACGACAAACAAACAGGA 540
Db 1054 GAACAGTAACGACAAACGACAGATAGTAACGACAATAACGACAAACAAACAGGA 1113
Qy 541 ACACAGATTAAAGCTTCAGAAAC 561
Db 1114 ACACAGATTAAAGCTTCAGAAAC 1134
```

## RESULT 2

```
KMAJ76 2198 bp DNA linear PLN 15-APR-2005
LOCUS KMAJ76
DEFINITION Kluyveromyces marxianus epg1 gene encoding endopolygalacturonase,
Partial.
ACCESSION AJ000076
VERSION AJ000076.1 GI:2597956
KEYWORDS endopolygalacturonase; epg1 gene.
SOURCE Kluyveromyces marxianus
ORGANISM Kluyveromyces marxianus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1
AUTHORS Siektele, R., Bartkeviute, D. and Sasnauskas, K.
TITLE Cloning, targeted disruption and heterologous expression of the
Kluyveromyces marxianus endopolygalacturonase gene (EPG1)
JOURNAL Yeast 15 (4), 311-322 (1999)
PUBMED 10206190
REFERENCE 2 (bases 1 to 2198)
AUTHORS Siektele, R.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1997) Siektele R., Gene engineering laboratory,
Institute of Biotechnology, V.Graicuno 8, 2028, LITHUANIAN
REPUBLIC
REMARK revised by author 04-NOV-97
FEATURES
source
1..2198
Location/Qualifiers
/organism="Kluyveromyces marxianus"
/mol_type="genomic DNA"
/strain="BMK Y-719"
/db_xref="taxon:4911"
/chromosome="VI"
338..>1655
/gene="epg1"
338..352
/gene="epg1"
573..>1655
/gene="epg1"
/EC_number="3.2.1.15"
/codon_start=1
/product="endopolygalacturonase"
/protein_id="CAA03900.1"
/db_xref="GI:2597957"
/db_xref="GOA:O13478"
/db_xref="InterPro:IPR000743"
gene
TATA_signal
338..352
CDS
573..>1655
```

## RESULT 3

```
AC116957.1
WPCOMMENT
Sequence split into 4 fragments LOCUS AC116957 Accession AC116957
Fragment Name Begin End
AC116957_0 1 110000
AC116957_1 100001 210000
AC116957_2 200001 310000
AC116957_3 300001 405682
```

Continuation (2 of 4) of AC116957 from base 100001 (AC116957 Dictyostelium discoideum c.

```
/db_xref="InterPro:IPR006626"
/db_xref="UniProt/TREMBL:O13478"
/translation="MLFSNLLIAAASALLAEASPLEKRDSCSLSGKTAGGGLSNCAT
VTNNVPEVPSGKTLDTGLPDGATVNFVQVTFGDEWVGPLVLSIGKIKVYVKGSGH
LIDGCGARWWDGDKGKKPKFMSLRLTNSDVGGGLQIKNTPIQAISSVNSCSDTVIH
DVTIINRDKDNLGHTDGFVGNVNTIENSHVYNQDDCIANVSGTGVYFKNYIC
SGIRKFGIIVGVDLNNVVDVTVYFENNQIVNSDNLGRIKTIQKATGSVNNVHFLSNTI
SGIRKFGIIVGVDLNNVVDVTVYFENNQIVNSDNLGRIKTIQKATGSVNNVHFLSNTI
TKWNVDIITGSSSFGSCTGIPSGGAPC"
573..657
/gene="epg1"
658..>1655
/gene="epg1"
/product="endopolygalacturonase"
/EC_number="3.2.1.15"
```

## sig\_peptide

## mat\_peptide

## ORIGIN

```
Query Match 99.8%; Score 560; DB 15; Length 2198;
Best Local Similarity 100.0%; Pred. No. 7.4e-75;
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 GCAGCCGTTTGGTTACAGTTCTCATGCAATCAGCCAGATTTTGCATAGTATTAACT 61
Db 1 GCAGCCGTTTGGTTACAGTTCTCATGCAATCAGCCAGATTTTGCATAGTATTAACT 60
Qy 62 TAGAATTAAGGCAACATCTTTGGATATGATGTAGAGTAAGTCGTTTCGAAACCATTTA 121
Db 61 TAGAATTAAGGCAACATCTTTGGATATGATGTAGAGTAAGTCGTTTCGAAACCATTTA 120
Qy 122 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 181
Db 121 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 180
Qy 182 TTGTTCTTAGTTTCACTACTATTATTATTATTATTATTATTATTATTATTATTATT 241
Db 181 TTGTTCTTAGTTTCACTACTATTATTATTATTATTATTATTATTATTATTATTATT 240
Qy 242 CAGCCTCCATACCGATTAGACAGGATCTCAACGCTGGGCTCCAGAGCTCACACATTATGC 301
Db 241 CAGCCTCCATACCGATTAGACAGGATCTCAACGCTGGGCTCCAGAGCTCACACATTATGC 300
Qy 302 TAAATAACTATCTACTGTAACTACAGGATACAGAAAAAACTATAAAAGAGCGAGGATAAAC 361
Db 301 TAAATAACTATCTACTGTAACTACAGGATACAGAAAAAACTATAAAAGAGCGAGGATAAAC 360
Qy 362 CACTCTCTTGTGAATCAGGATCAGTAGGTAACCTATAAAACCTTCTTTTCTCTCAAAA 421
Db 361 CACTCTCTTGTGAATCAGGATCAGTAGGTAACCTATAAAACCTTCTTTTCTCTCAAAA 420
Qy 422 TATCAATAACAGTAGTATCAACAGATATCGAATAATACCTACTACTACACAGTAGG 481
Db 421 TATCAATAACAGTAGTATCAACAGATATCGAATAATACCTACTACTACACAGTAGG 480
Qy 482 AACAGTAACGACAAACGACAGATAGTAACGACAATAACGACAAACAAACAGGAA 541
Db 481 AACAGTAACGACAAACGACAGATAGTAACGACAATAACGACAAACAAACAGGAA 540
Qy 542 CACAGATTAAAGCTCAGAAAC 561
Db 541 CACAGATTAAAGCTCAGAAAC 560
```





```

/translaton="MNVDIIVADIIIESENDOENDKMSOSPTLKEDHNKIKE
PEEIEENLKEISNLIKESKNEIETLSKQIFSIIVKILRLDYKATDCTKVHDD
PWEIYKVIKIRHNOVQNTLOKLESEGSLLKESQYEGFVNTKDGYNRWLPS
AHEVYLDQFNQWKNKISPLERDNYGRWISFIPNNSGECALPGSKGKIYKLANG
FYDIRPAWIKVEQKELGVDFVFNPSKQVFNKSPKMPKTELRIYEAHVGSSEL
PEIYTSFKDTLPVWKELGNCIQLMAVMEHAYIASFGYQVNFASISRFCTPEE
LKEMIDKAHEMGLVFLDVHSHASKNVLDGLDQTDGDDHYPHSGGRNHELWDSRL
FYNGWYEMFLLSNLRYVDEYHFDGFGVGLQVDTSMIYTHGLSPACSYDDYFGADV
EDALNYLTANVMLHTLNPISVTTAAEVTGLATLCRPESEGGGDFDVLAMGIDPKKI
ELVKEKEDNMGTIAHMLSNLRYKERNIAEASHDOSLVGDKITLAFWLMKMEYNTN
MSVTEETPIDRMSHLMIRLITSSIGGGGYLNFMGNEFGEHPEWDFPREGNNSL
HAARRLDYXNPLRYKQLRDFDIAMNKAQEFRLSSDFAYISLKHEDDKIIVFER
ASLIPFNHPSKFSDFYIGSGVFGKEINVLDSDRKEFGFVGRIGKDNVHYTEDKFW
HDRKSLIYIPSRCLVLKVD"
Join(15053..15120,15221..15743)
/note="GeneID exon scores (in order of location ranges):
-0.56, 37.71 - GSCJ_ID dd_01882"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA08769.2"
/db_xref="GI:28829251"
/translaton="MGQKSLGELQYPENMKELFNKIEDCKYSYDTKDDTFLSQYND
NNKQKSNDSISGNKNSKEQSSSTTTTTRRNNSKPEPIKTSNGVTMGSLSL
LNCSTNDSDDDDDEEDDEEDNNSFENNPLTEBEKOKIIOEKYKQOQKK
KLKANKKIIRKDPYWFGEISTPSLKOSQSYFKQKG"
complement(join(16609..16729,16860..16934,17034..17080,
17169..17177))
/note="GeneID exon scores (in order of location ranges):
6.96, 13.18, -2.05, 0.27 - GSCJ_ID dd_01883"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA051794.1"
/db_xref="GI:28829252"
/translaton="MAGWTIEPGKHGGITWAPRSTTVHVGVGGGRVRMQAG
ASPEDNDVNGETSFSGFSGRLNVTNIGSTLKVWTA"
complement(join(19502..21007)
/note="GeneID exon scores (in order of location ranges):
108.08 - GSCJ_ID dd_01875"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Futa (Fragment)"
/protein_id="AA051795.1"
/db_xref="GI:28829253"
/translaton="WKSLSNSYILLKIVISTCLLLINIALIYEDPNLIDSNCKTT
TTITQNNKNGLENLNSKKLLILDYKLENDTYMKSSECTSIINQPCGT
TDSLANSDEGLIFERTYKNRYSYTSVHRISERKQLFQWMSFSPYPLENDT
HLSUNYVTIGYPLDYKYGAGVDEEYEDKIOSSHITYPYGVFPFGSDHYSYL
YGNVLLRSLOKILKRYDFVAIENSRCEDYVEKLWPLSVGTPIYLGATNIDE
FLPHDPLINTANFNSINHLWKYVDVGLNETLREKLAWIKPLPKFKNLLNOSLN
NQLFCSCIKLVNLDNNINYPPLSKPKQCEGKFTPLPYTEEKERGLRLQK
QOEKLELQNNNDYNDNEEEDNGSDNGSDNGSDNDSN1"
complement(join(21960..26819,26946..27251))
/note="GeneID exon scores (in order of location ranges):
530.44, 21.78 - GSCJ_ID dd_03196"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Nucleotide exchange factor RasGEF R"
/protein_id="AA051796.1"
/db_xref="GI:28829254"
/translaton="MKSEETANLGPQBYFNYLEGINKKIIILKTVANIDTAXDTNUL
QNILLESALERNILLYVDKPLSYDQIVYHSHFKEVVKIPQVQCQDLKAAQ
ALLNSQDLYTTSNPAKRGAEIKTIKNATSNFELKLYNFDQLDLEDLVHQHQ
QKEQEQKQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ
EERKNDKEKEKELKQQLAEQKQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ
QRSDPIIPTITLTSQDLSGLSPKSPGSPGSKFTESAERVANFDRTGTAII
LNNKIKESLGSDEQAEPIVTAARIISENIAIISKELRYKTLGNLSHFVSQVQ
ARNAIKNSQFQYDQLELALEKENDTMRKIIYVSKISKSNLQAFGVDESSLRS
PVVSPKISPSFTSSSERILSHKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QQNPQPLQLLNLQSLSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
TSTSTSPSSSSLSLSDQKQKREKQHVIPKVDKSVKVAQOOOQOQOQOQSTPR
NNSSVTAATTTTSTTASITNVNTIIPNFPNKPQLPATPTTGTPTSTPTPTST
SONDKQNNNNKENFVDKQTKGLFSKFEVHKRTPLPGFDSSSSSSSSNNNTTNS

```

```

SHSSTNSPMTSPGVESPKITKSPSONNILLVSLDSDIGSDNQOQOQDKLTITTTTIT
TINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
RTTIGLVSKSKSCVVVEYFKSKLFOHESAKQLLSIIISWNPFSFEKFPQONSEVIS
NIIIEQIGNVINKYHDEVSESSSSSSSSSTTSPNNINTPDCSPILNSNSKDSHNLSSI

Query Match      17.2%; Score 96.6; DB 2; Length 333321;
Best Local Similarity 69.1%; Pred. No. 5.2e-06;
Matches 132; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 40 AGATTTTGCAATAGTATTAACCTTAGAATTAAGGCAACATCTTTGGATATGATGAGT 99
Db 287902 AAATTTATCAATAGTAAAAATCTAAATTTTCAATTTCTTTAAATAAATTTTGATATT 287961

Qy 100 AAGTCGTTCCAAACCACTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 159
Db 287962 ATTTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 288021

Qy 160 ATTAGTATTATTCGAATTTGTTTCTTCTTAGTTTCACTACTATTATTATTCATATTCAT 219
Db 288022 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 288081

Qy 220 GTTATTCACAT 230
Db 288082 ATTATTATTAT 288092

RESULT 5
AC116960 182870 bp DNA linear INV 12-MAR-2003
LOCUS Dictyostelium discoideum chromosome 2 map
complement(1004496-821614) strain AX4, complete sequence.
AC116960
VERSION AC116960.2 GI:28850330
KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 182870)
AUTHORS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P.,
Lehmann, R., Baumgart, C., Parra, G., Aprill, J. F., Guigo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Noegel, A. A.
Sequence and analysis of chromosome 2 of Dictyostelium discoideum
Nature 418 (6893), 79-85 (2002)
JOURNAL 12097910
PUBMED The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 182870)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 182870)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 182870)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Mar 5, 2003 this sequence version replaced gi:19920015.
CDS predictions from GeneID do not necessarily reflect true genes.
Further information is available from IMB Jena, Department of
Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml)
Funding Agency : Deutsche Forschungsgemeinschaft (DFG).
Location/Qualifiers
1. .182870
/organism="Dictyostelium discoideum"

```





/note="GeneID exon scores (in order of location ranges):  
22.72, 7.57 - GSCJ\_ID dd\_00341"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AA051289.1"  
/db\_xref="GI:28828690"  
/translation="MTPEYKHDKSGDTSNGYSLIVNECFAPDRPFMVFEFNHDSN  
ATINQVHDKCSKIYNEITFCINECYOAPYVWNSGAEANFVKISISTPNYIPO  
YGHROTTFPGDINCQSNPWFYVATNETKLYNDEEVIIFCIDIQNPYKTLNCPCEED  
YSSLSCVSSSDVDKDPFGTEVC"  
join(14195..14991,15224..15551)  
/note="GeneID exon scores (in order of location ranges):  
28.36, 11.51 - GSCJ\_ID dd\_00325"  
/codon\_start=1  
/product="similar to Plasmodium falciparum. Hypothetical  
protein"  
/protein\_id="AA033197.2"  
/db\_xref="GI:28828691"  
/translation="MTIITPLFYIQKFIIRLISHNIDQNRLIDICSSGSSSVTSY  
RNEKRIKTLILMSLVSEWFKTISNFLTIVSDFNVDODODOOOQHSYIIEKENIET  
LKLVDHKEKSVYKLSGCKYIKIELEVNKKIMKLSKHOGGKGNFKNLTIRVNY  
ONEFRIIDSLYOCEQNLIDICEFSLVRSIQOQENNNNNNNNNNNNNNNNNNN  
NIETLIIDNAGCDFEITKHSNSIKLIVSKAFYPLFOLVFRCHNEINNNNN  
NNNNNNNNNFEFEIPKRNINIKLVYDKCIYQDLTKTKILDTSEIEIFISFLS  
LPKQKTFETNIPSEFELDEIIEKQNTIPLHLYLGNQAD"  
complement(join(15648..16516,16614..16767))  
/note="GeneID exon scores (in order of location ranges):  
100.76, 14.23 - GSCJ\_ID dd\_00327"  
/codon\_start=1  
/product="similar to Ralstonia solanacearum (Pseudomonas  
solanacearum). Probable alcohol dehydrogenase (ADH-HT)  
oxidoreductase protein (EC 1.1.1.1)"  
/protein\_id="AA051290.1"  
/db\_xref="GI:28828692"  
/translation="MTKTKMAAFTQKNGKIEIIELPPEPTQGWIRIKVHACVQC  
GENVCKWGNMSPFRPGHVEYIDKLGVCNEEYKIGFVGVFGNGCQKCE  
TCLNEHKKKNTCVGTGYYGAYEIMAPISALVKIPKGMDFLEAAPLLCAGVTY  
NSFRHDKYGLSVGTGIGLGHVAFIOCKRMGYQVLIAMSGNSKEOLSKELGADY  
VDMKQDIOETISGTSVKILATPIASTVQGLSGLINGKLVILAFEPHPHADS  
LTMIGGSKVWASGDRNDSLTINFARNNNISLVNTPLEKANEAFINQAKFR  
HVKUL"  
complement(join(17462..17749,17849..18038,18121..18263))  
/note="GeneID exon scores (in order of location ranges):  
29.13, 12.86, 3.70 - GSCJ\_ID dd\_00329"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AA051291.1"  
/db\_xref="GI:28828693"  
/translation="MEDIEINLSKDKKEFTQLLESTTPTDKLAFNFORVLIN  
OETSNFKFLQDNTEENYKGCOSITTLSSIGEEIIIOESLSTNPMSKILSEVR  
ENKTKFLTAQNVKVEIDNSQSLTSKNKNEKPAAGSCFYEHIEKCESQINENRLK  
INDCITINEKLEILKLDNLEFAGLNTDKENOSHHSHHHHK"  
join(18932..18972,19079..19364,19446..19715)  
/note="GeneID exon scores (in order of location ranges):  
5.11, 32.60, 18.52 - GSCJ\_ID dd\_00333"  
/codon\_start=1  
/product="similar to Xenopus laevis (African clawed frog).  
Glutathione S-transferase"  
/protein\_id="AA051292.1"  
/db\_xref="GI:28828694"  
/translation="MSTSSVPSLTVFQGRGLGQPSRVLLSYLGIPYENITVTEISDAL  
RATLPYGOLPIYRGDFVLTOSSTIARYIAKHMPMGKLEEFVDOIVTAIHADIF  
PAFNPVPEKQLKLEFKFGSKQETGLVSGSVTLADLYVVGDFYIRFGEAA  
LSSELSDSKYPIAKELKFKFSGNCGVAKIYKERTPK"  
complement(join(19893..20698,20780..20921))  
/note="GeneID exon scores (in order of location ranges):  
107.38, 10.99 - GSCJ\_ID dd\_00336"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AA051293.1"  
/db\_xref="GI:28828695"  
/translation="MVVVTESIVNGLKLTLDVSOPIAKRPSRLFTLRNLNGLPICDA  
MASALNDSKALLRHEIAYCLQMEDEYALKLIDLVKNSDEHPVNRHEAEALGAIGS

ESAKHTKEYSNDPVREYSETCOLALSREVMYKKNKPEEDKMYMSVDPAPPLKKG  
VSRDLRSKFLDSNLDIFNRYRALPSLRDIDGDEQSVLACDLGKQSSALLREHAFV  
LQLOHRAVDIPLTTCVLDSESEAMVRHEAEALGAIATETIPLLEKLLQKKEPIVS  
ESCAVDVTEYFNNTESFQVADGKIKLEKLVQOQK"  
complement(join(21187..24684,24816..24929))  
/note="GeneID exon scores (in order of location ranges):  
286.22, 0.86 - GSCJ\_ID dd\_00338"  
/codon\_start=1  
/product="similar to Dictyostelium discoideum (Slime  
mold). Putative RNaseH1"  
/protein\_id="AA033200.2"  
/db\_xref="GI:28828696"  
/translation="MIVNSRLLNEFLQYKIGPTPVFSTKNEGLDHPFFTTTCIYL  
GHEFVSTKSKKEAEADCSQLVLDLFERQTOQQPQQQFQNPILPKLQLLNQO

Query Match 17.1%; Score 96; DB 2; Length 110000;  
Best Local Similarity 70.1%; Pred. No. 8.9e-06;  
Matches 129; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 42 ATTTTGCATAGTAACTTAGAATTAAGCAACATCTTGGATATGCATGTAGATAA 101  
Db 97803 ATTTTAACTATTATCCAAATTTAAATTAATTAATTAATTAATTAATTAAT 97744  
QY 102 GTCGTTCCGAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 161  
Db 97743 GTTGAATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 97684  
QY 162 TAGTATTATTGAATTTGTTATGTTCTTACTACTACTACTACTACTACTACTACTACT 221  
Db 97683 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 97624  
QY 222 TATT 225  
Db 97623 TATT 97620

RESULT 7  
CR293524/c  
LOCUS  
DEFINITION  
Zebrafish DNA sequence from clone DKEY-57M7 in linkage group 18,  
complete sequence.  
ACCESSION  
CR293524  
VERSION  
CR293524.10 GI:54261125  
KEYWORDS  
HTG  
SOURCE  
Danio rerio (zebrafish)  
ORGANISM  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
REFERENCE  
1 (bases 1 to 190196)  
Wood, J.  
Direct Submission  
Submitted (15-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Oct 15, 2004 this sequence version replaced gi:53628335.  
COMMENT  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Web site: <http://www.sanger.ac.uk>  
Contact: [zfsh-help@sanger.ac.uk](mailto:zfsh-help@sanger.ac.uk)  
-----  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least



[illegible]



|     |  |   |       |
|-----|--|---|-------|
| CDS | EDIEENFDL"   | complement (join(17257..17777,17920..18049))              |       |
|     | /note="GeneID exon scores (in order of location ranges): | 49.00, 9.03 - GSCJ_ID dd_00530"                           |       |
|     | /codon_start=1   |   |       |
|     | /product="similar to Arabidopsis thaliana (Mouse-ear     | creas). Apospory-associated protein C (Hypothetical 35.4  |       |
|     | kDa protein)"  | /protein_id="AA050943.1"                                  |       |
|     | /db_xref="GI:28828269"                                   |   |       |
|     | /translation="MYTELOQLSNKNGKQVIOVLKGNNDLSKLVKLTSSSCEIYIH |   |       |
|     | AHITSFKTNSVEHLFMSEKSGQDMKAIKRGVPLIHPQNGKIQHPGFARNCDMEIH  |   |       |
|     | KALVDEKTSQVLEFLKRNNESEKWNKCNITSGKTIINFTIYKVLHPFLDLEF     |   |       |
|     | DVKNITSPSNDNILEFQAFHYQISNNVYVILGDRGYIDKMGHKTKE"          |   |       |
|     | join(18919..19049,19137..21735)                          |   |       |
|     | /note="GeneID exon scores (in order of location ranges): | -2.24, 179.96 - GSCJ_ID dd_00531"                         |       |
|     | /codon_start=1   |   |       |
|     | /product="hypothetical protein"                          |   |       |
|     | /protein_id="AAJ93574.2"                                 |   |       |
|     | /db_xref="GI:28828270"                                   |   |       |
|     | /translation="MGTFKKLFIFFLIINIILCYCGCEFSIGENKYNFTAPFG    |   |       |
|     | EYTSQGTNTVFLNCTNQISKPTTSACIKDSKNTDIGOQVPAFLPKGILLT       |   |       |
|     | YKQAYPCNTVGLGVSTNTICNSDTSNFTLDKIQQFSICYSKIFTKLACVCPG     |   |       |
|     | DCSSHGICNTIKCEBNSVASSGNSKELKMIISTISITIIIGDVMHSGDFSNPLE   |   |       |
|     | IKNNLKINPDGKCVINNTNTOCTADKYGNGIEVEFSSGLSTVIVPNFKYI       |   |       |
|     | EIPCFNCSYGVNCKILGTCNHSINGTGCELIKTHSIEPTTGITIIIG          |   |       |
|     | DTFMIINGTITVLPNLLIKIGDDDDSSLCQDKILNDNLQCTVLPKQGGKNN      |   |       |
|     | TISGLSDPNYQLFVFNKTCFNCYPHGCDLFTSCDSONFGNGCKSNLVI         |   |       |
|     | ESIDPTDEGLINVGFGPIGTPISIKIGDILCNLLIINQSLQCVVPKGFKN       |   |       |
|     | VFISDRSLSPNGMSTFVIVTNAPKQCTNGALNQGVCASQICHPHMGVDCQ       |   |       |
|     | SOXIIIPQPPKNSPVEPTSPKNSITGGGSGTSDIDENENIKNVFKSIIAI       |   |       |
|     | IKSELDFQSNKTPFSPSWIKELTPSKYQVNTIKVTTNDGGGVSPSSSSSL       |   |       |
|     | NESETNISVLENFESLITITFANELINMPSSIKVITBISKYKSSKLNLQLVWMA   |   |       |
|     | ELSNKTKDLCNNFETSGNSNYIKLOVDNHSLYGRFKRAIDISYPAIGNVQ       |   |       |
|     | LDEMRPLKNSQSYGVIVPFKSIIDPDFSVLIDSSKLENSNSNSICLIQD        |   |       |
|     | QLSLGTSQIAGIITGSIAPLCSITAMVIYSIHKSNDKLSFKITLNLNK"        |   |       |
|     | complement (21934..22398)                                |   |       |
|     | /note="GeneID exon scores (in order of location ranges): | 21.03 - GSCJ_ID dd_00532"                                 |       |
|     | /codon_start=1   |   |       |
|     | /product="hypothetical protein"                          |   |       |
|     | /protein_id="AA050944.1"                                 |   |       |
|     | /db_xref="GI:28828271"                                   |   |       |
|     | /translation="MDNLKSKPLSYNNNNNNNNNNNNNNNNNNNNNNNNNNNNF   |   |       |
|     | KYVSPSLEDIGYQYIDKEDYVIOQVYIRQNSDDEEDFNNNNVEDSIKIS        |   |       |
|     | LIQKSFINKKNIITIIIVLMIASVGLLAWGVFDQIHNNNSKDE"             |   |       |
|     | complement (22774..23184)                                |   |       |
|     | /note="GeneID exon scores (in order of location ranges): | 28.48 - GSCJ_ID dd_03186"                                 |       |
|     | /codon_start=1   |   |       |
|     | /product="hypothetical protein"                          |   |       |
|     | /protein_id="AA050945.1"                                 |   |       |
|     | /db_xref="GI:28828272"                                   |   |       |
|     | /translation="MWSIDFNSSIVQNQNNFSPATGNDTITSTISLEKPSDNIYA  |   |       |
|     | NEELPFDPDSVNN      |   |       |
|     | VNGLSRKTFRIISTIIISLIVCIIVLVIVFGKN"                       |   |       |
|     | complement (23397..23567)                                |   |       |
|     | Query Match  | 16.7%; Score 93.8; DB 2; Length 36148;                    |       |
|     | Best Local Similarity                                    | 57.7%; Pred. No. 2.7e-05;                                 |       |
|     | Matches  | 206; Conservative 0; Mismatches 147; Indels 4; Gaps 2;    |       |
| Qy  | 77   | ATCTTTGGATGATGATGAGTAGGTCGTTCGGAACCATATATATATATATATTA     | 136   |
|     |  |   |       |
| Db  | 24855  | ATGAGTGTATATTTGTGAGAGAAATCGTGTTCATATATATATATATATTA        | 24796 |
|     |  |   |       |
| Qy  | 137  | TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA    | 196   |
|     |  |   |       |
| Db  | 24795  | TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA    | 24736 |
|     |  |   |       |
| Qy  | 197  | CTACTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA | 256   |
|     |  |   |       |



```
/db_xref="GI:28828451"
/translation="SONQSEDOKLLFGLYSLKAFITSSPKKIDDKTGPHCYKTSY
KLHYETLSICKFIIMDPNVDLDDLLKLYSIQIFVEYIKNPYKHGTGYKCDIFI
NQLNLYLKQMPSESS"
CDS
complement(join(3621..3656,3814..3876,3996..4504,
4659..4986))
/notes="GeneID exon scores (in order of location ranges):
0.03, 5.79, 47.43, 25.62 - GSCJ_ID dd_00208"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51035.1"
/db_xref="GI:28828408"
/translation="MKFIVSIFIIQIIFINLINGQIVRGDQLQGVYENGTMILGS
PNVNTGDQNMNTKOLSNWTDHRYNDWDFIRPHRNNQSGFEATDFPRVDE
FSTIKSEVGEYTVMAEIPYSKNVTSFTSLVDVPCQETLTGWSVVVTDKACQVTD
ITATYQKVENYNSVLEPILPPKHIFPESRTFVSFYTFIIPQLNGALVDAWDGD
RSTICNCDSCIDNOCATKDFDDMGEGCOLKFFFLAWAGVDKSGKNCVSLNKIPS
KFOKYSATAIKNGTGLVSDFFYKINDNNNPNTA"
CDS
5802..5957
/notes="GeneID exon scores (in order of location ranges):
4.32 - GSCJ_ID dd_00209"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51036.1"
/db_xref="GI:28828409"
/translation="MRKVSYKYPENHKQSNLENCLTFYQMIQTDPDRSIFKLATSNLL
ISNIKQ"
CDS
complement(join(6544..21168,21250..21832,22134..22616,
22677..23743))
/notes="GeneID exon scores (in order of location ranges):
1211.90, 27.13, 26.28, 60.75 - GSCJ_ID dd_01256"
/pseudo
/codon_start=1
join(24858..25024,25139..25410,25452..25831)
/notes="GeneID exon scores (in order of location ranges):
16.28, 24.00, 42.08 - GSCJ_ID dd_00213"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL96709.2"
/db_xref="GI:28828410"
/translation="MSTQGLRFLSALRKAAGMSTPGAGNSVSNLEDFNDVAKSVH
KTDSTNLLVGDSPGDKILLSTGKINELIEHFDSTSMYAILRVVDHVDGPT
DRFVITFTGDNVAKIEKAKYSVNTKVTLLGHVNVETIASELKEQVTSDDPNMM
LIGYEDKRLGVGTGLOSLVSHLDQNVYGVATTDVIDGIVOPKHAQITFVG
VSVSPVWGRLTHRGKIDFFSPHVSFLPSNPQDLTDEILAQVQLKGKMK"
CDS
complement(join(26042..27140,27220..27341,27450..27770,
27854..28531,28636..28935))
/notes="GeneID exon scores (in order of location ranges):
109.13, 7.33, 29.97, 57.67, 11.17 - GSCJ_ID dd_00215"
/codon_start=1
/product="similar to Arabidopsis thaliana (Mouse-ear
cress). Hypothetical 80.1 kDa protein"
/protein_id="AAL96711.2"
/db_xref="GI:28828411"
/translation="WTTNTIINNGVSFLNNSNIPTIHPSPNPVIDCRDGTWAY
IYLVDESYYQQHNPSPKPSLPLPKLYSTPNKTPYPQFOQNNENISIVKEIY
NYNCFEIKDCDFCDVDSIKYARPHNLTLKMGCKSIHARQISNTVFQLEKCR
EQNKUQTANNAIITDTSIKETSTTTTTTTTTTTTATTTTQPIYCVSIQRN
IFYIIGHLSQYCSQPOYIIDFPVSCVQDKHSFTLLGSGCGKSTLTALLAS
RIGFAVSTDNIRQLLRKFIQRSPILWASTHAGEIISNPSLHKEKILQGYEAO
NEMFNKLDILGHYKREKESLIVGVDLTKLILRVKKHPSGCPFLMYISNEAKH
BFAIRSKYMTLDPHONKYTKYKNIIRINDHLCHGADHMIPOIDNTSIDRSLATIH
GTIPACLRKVCQGSYNYHETDKMMLYNOYEIQHQFWSKGMRLLIQKKTSPH
GNNVTGDVNNNSNNNNNDGNSDNNNDNNNNNNNNNNNNNNNNNNNNNNNN
DNNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDN
YKNNNSNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDN
HNYTQNNLNNNNKSDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNND
IQEQQNNLNNNNKSDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDNND
NHNNGNOSDNDDSDSDSDSDGGERIDYVLCGSLGS"
join(29669..30019,30109..30626,30737..30996,31116..31276,
31353..32132)
/notes="GeneID exon scores (in order of location ranges):
29.16, 56.32, 27.47, 16.22, 92.26 - GSCJ_ID dd_00200"
Query Match
```

```
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL96712.2"
/db_xref="GI:28828412"
/translation="MSVPSSVIEHNLVPEKIPVYRLNARESFLNLTEKEQLYAHHSV
ACWGSKICLGQTSIESGFIENLQNLKSTVNPVNISEESYSLLSYAATFY
GMNGYLSGDSKFIPIRSKEKQLIINKNDKNVNEYMGKSELMSLDKOVRELGI
DNGISTVYSPNITKVEIKQKFDKSDKISIPVNTLFPVSENYNLLASASTSTPT
VSHQFDGYTINIYGDNNKLTUVVDNLKALPAAENQTNMLKXIIISFYSGSIDD
HDSQRWIKDISPAVENTIGIESIRDPYGVGEWGFVSVNKEMSUKPGKLTUNA
TTFLSKLPMDSFEKENKPDFTLEVTATTPAGINLNDNDIRQTEGFKNVLG
NVIAAGDEYTVFIOESQKLFNELSTAFELQVGIHELHYGSGKLCGCIYLSPEKILE
KVGVEINLTNKPIDPKTEVYKFGTYSFKSLGSPMECEKAKCCGCIYLSPEKILE
LFGFTDPKAEADVIVNMLINARAGVCALEFVPPSEGAQKRWQAQMOARCYLITFE
LBSGLVTLDKTADDVIVKLDKSKIRGIVKAVGDFLRLMYKATANIIDASIKLFBY
THVNEBFLARDIVLAKKPRKVFQAHTILNSNGKVCLOQDFDQGMIDSMITRFG
KDDSDML"
CDS
complement(join(32452..32600,32953..33619))
/notes="GeneID exon scores (in order of location ranges):
-0.20, 32.06 - GSCJ_ID dd_00202"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL96713.2"
/db_xref="GI:28828413"
/translation="MTEVTYQIQDQIKQFNENPIQKTNKYQYNSDNENSSIVYLK
NPQNPNYNNKICNYKRSLSSSSQFQNEKETRKNKAQFVENDFENGFGNISML
NNSIPYKCEDCNGNEPPOKLSQDILPFPPEPLTHEINKVHIELETAGVDKD
DVVDLTNLTIVAKKSVPLPQNMCFKPKHEKSGVYKRVLSFNSNTVOKDTIKA
RYVNDSPKAYMACLMYGVDFHINTGLDYCPGELESOKLQNDYFQRYQI"
complement(join(34387..34623,34721..34883,34986..35770,
35912..36055,36141..37055,37130..37763,37866..37951))
/notes="GeneID exon scores (in order of location ranges):
13.33, 11.81, 81.94, 11.24, 89.27, 61.07, 13.05 - GSCJ_ID
dd_00204"
/codon_start=1
/product="similar to Mus musculus (Mouse). Adult male
liver tumor cDNA, RIKEN full-length enriched library,
clone:C730036A03 product:ESOPHAGEAL CANCER ASSOCIATED
PROTEIN homolog"
/protein_id="AAO51037.1"
/db_xref="GI:28828414"
/translation="MAERQSASSTPSSPPQQOQTPQPPQYLLNSKNFKVKVING
REVERLNNPLSKHDPLESETATWGGLKIKENHLTVYDENFMPTDVLKSILOQYT
TDDNPIQVSPMSTGTSKIKIPINRLNKILEELQDKEDSKSTQSPQDIDMLETL
HSELQAAEERVRSKIAIQAKLLSDLSIKFIPSKFVIAIILDTFGNLVDRI
KRLQSSKEKHEILLKEQAKETRNWFIASIRELLPRLFVIEISILCKYEFQGD
VNTPEQVINRISIRIGNPLVANYIRAYLTRSFDCPEYKFLQLKDFVFTQ
KSYEKSYLENTLSMYRITLTDYMLGYSPLSLWLLQCLAHKATPETLEVLFRSK
NLLNHTIISFPPEYICSNMTFSNFIKDADTLSPYKQVSTFGVNLVLPQPKNQ
ILSLNDVWVYTNFENIKDYLSAEVIEYVLTCHSEKETDVLKDLIRHLIPKGY
ETIHLQSVIKFTIHISDFGLVFNFLPDLDFNGESQKLSRSTLEALSTSKV
MTSDPILNTFLTYKALHDSLSLSFQDFQVTVLVVNCINKIDFGRDVBKQLNFY
VECTQTFNFDGVKNRLYGVCEICEKTLNVKGHTKPTTSFIRACVAYCFITPSI
YDIFKMLYLVSSVLAQNALQADALLKAAITFIOEIPFLFQKSTQEDTMSWA
VYSDPILNLTVPCHPSGPFYVAKYKVIQWQESSSTAKSKFLQLLCCSSWA
QTSPLPHIEKVESNDQLTPEPNTTELTFEYNSLIKILYDLNLLKDEPNLTQKV
GIICLDLNLNVLNGLNSLTASLFDNNAKKIIPSTCFNEITLKNLAFIGLIE
SKMGQDIFNKLSSQQQ"
CDS
complement(join(38390..42839,42945..43705))
/notes="GeneID exon scores (in order of location ranges):
443.99, 41.08 - GSCJ_ID dd_00207"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Histidine kinase Dhk1"
/protein_id="AAO51038.1"
/db_xref="GI:28828415"
/translation="WIVYLKLEKLNQALWDFINKRVLWGNOSCKYKNIETSILF
NNSNLINDIRCKSKHEHIDIESNENGEILKOMIFKYSSIHISMDARQLYLDTSIN
LLVNDQSHLYDEGRKIFILVEAIEIKFQSPSSSENHNINNNQNSVNNSS
NKQYNPEPNSMGSWENWVQNDTTKASNQFIKGIENDFNKLNFDNFKUGIQE
16.7%; Score 93.6; DB 2; Length 132254;
```





|                      |  |
|----------------------|--|
| REFERENCE<br>AUTHORS | 1 (bases 1 to 146570)<br>Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,<br>Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,<br>Tongal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and<br>Neugebauer,A.  |
| TITLE                | Sequence and analysis of chromosome 2 of Dictyostelium discoideum  |
| JOURNAL              | Nature 418 (6893), 79-85 (2002)  |
| PUBMED               | 12097910   |
| REMARK               | The Dictyostelium Genome Sequencing Consortium   |
| REFERENCE<br>AUTHORS | 2 (bases 1 to 146570)<br>Baumgart,C.   |
| TITLE                | Direct Submission  |
| JOURNAL              | Submitted (06-APR-2002) Genome Analysis, Institute of Molecular<br>Biotechnology, Beutenbergstr. 11, Jena 07745, Germany   |
| REFERENCE<br>AUTHORS | 3 (bases 1 to 146570)<br>Baumgart,C.   |
| TITLE                | Direct Submission  |
| JOURNAL              | Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular<br>Biotechnology, Beutenbergstr. 11, Jena 07745, Germany   |
| REFERENCE<br>AUTHORS | 4 (bases 1 to 146570)<br>Baumgart,C.   |
| TITLE                | Direct Submission  |
| JOURNAL              | Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular<br>Biotechnology, Beutenbergstr. 11, Jena 07745, Germany   |
| REFERENCE<br>AUTHORS | 5 (bases 1 to 146570)<br>Baumgart,C.   |
| TITLE                | Direct Submission  |
| JOURNAL              | Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular<br>Biotechnology, Beutenbergstr. 11, Jena 07745, Germany   |
| COMMENT              | On or before Feb 21, 2004 this sequence version replaced<br>gi:28828883, gi:28829496.<br>CDS predictions from Genes to Genomes do not necessarily reflect true genes.<br>Further information is available from IMB Jena, Department of<br>Genome Analysis<br>( <a href="http://genome.imb-jena.de/dictyostelium/">http://genome.imb-jena.de/dictyostelium/</a> )<br>and the University Cologne, Institute for Biochemistry I<br>( <a href="http://www.uni-koeln.de/dictyostelium/project.shtml">http://www.uni-koeln.de/dictyostelium/project.shtml</a> )<br>Funding<br>Agency : Deutsche Forschungsgemeinschaft (DFG).<br>Location/Qualifiers<br>1. .146570<br>/organism="dictyostelium discoideum"<br>/mol_type="genomic DNA"<br>/strain="AX4"<br>/db_xref="taxon:44689"<br>/chromosomes="2"<br>/map="332358-3470138"<br>complement(join(<3. .86,436. .603,692. .704))<br>/notes="GeneID exon scores (in order of location ranges):<br>1.46, 2.66, -0.45 - GSCJ_ID dd_03180"<br>/codon_start=1<br>/product="hypothetical protein"<br>/protein_id="AAS38882.1"<br>/db_xref="GI:4273398"<br>/translation="MTILGISISNVKSIKSNNSLSISNSLSISQSMNSIQQCGG"<br>SISNPFLGIGRNPGESMVCDSG"<br>complement(join(1507. .1763,1872. .1884))<br>/notes="GeneID exon scores (in order of location ranges):<br>19.50, 3.22 - GSCJ_ID dd_01276"<br>/codon_start=1<br>/product="hypothetical protein"<br>/protein_id="AAS38835.1"<br>/db_xref="GI:42733927"<br>/translation="MTILGISISNVKSIKSNNSLSISNSLSISQSMNSIQQCGG"<br>GNGLLGCGVGLVGVLTGVGTIGSVLGVGSLTGGSNCCGN"<br>join(2611. .2623,2691. .2947)<br>/notes="GeneID exon scores (in order of location ranges):<br>3.18, 23.32 - GSCJ_ID dd_01275"<br>/codon_start=1<br>/product="hypothetical protein"<br>/protein_id="AAS38836.1"<br>/db_xref="GI:42733928"<br>/translation="MTILGISISNVKSIKSNNSLSISNSLSISQSMNSIQQCGG" |
| FEATURES<br>source   |  |
| CDS                  |  |
| CDS                  |  |
| CDS                  |  |

|                       |                 |                    |           |                |  |
|-----------------------|-----------------|--------------------|-----------|----------------|--|
| Query Match           | 16.5%;          | Score 92.6;        | DB 2;     | Length 125623; |  |
| Best Local Similarity | 64.9;           | Pred. No. 2.8e-05; |           |                |  |
| Matches 137;          | Conservative 0; | Mismatches 74;     | Indels 0; | Gaps 0;        |  |

  

|    |      |  |      |
|----|------|--|------|
| Qy | 15   | TTACAGTTCTCATGCACAAATCAGCCAGATTTTGGCAATAGTATTAACTTAGAATTAAGGCA | 74   |
| Db | 6526 | TTTTATTTCCAATTAAATTTGCAATTCCTTTTAACTATAATGGATCCTTAATTAAATTA    | 6585 |
| Qy | 75   | ACATCTTTGGATATGCATGTAGAGTAGTCGTTCCGAAACCATTAATTAATTAATTAATAT   | 134  |
| Db | 6586 | CATAAAATGTCATCATCTTTTTTCAATATTTTTCATTTAAATTAATTAATTAATTAATAT   | 6645 |
| Qy | 135  | TATTATTATTATTATTATTATTATTATTAGTATTATTGAAATTCGTTATTCTTCTTAGTTT  | 194  |
| Db | 6646 | TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT      | 6705 |
| Qy | 195  | CACTACTATTATTATTTCATATTCATGTATT                                | 225  |
| Db | 6706 | TACTACTATTATTACTATTATTATTATTGATT                               | 6736 |

  

|            |   |           |     |        |                 |
|------------|---|-----------|-----|--------|-----------------|
| RESULT 14  |   |           |     |        |                 |
| AC117076   |   | 146570 bp | DNA | linear | INV 21-FEB-2004 |
| LOCUS      | Dictyostelium discoideum chromosome 2 map 3323568-3470138 |           |     |        | strain          |
| DEFINITION | AX4, complete sequence.                                   |           |     |        |                 |
| ACCESSION  | AC117076  |           |     |        |                 |
| VERSION    | AC117076.3  |           |     |        |                 |
| KEYWORDS   | HTG.  |           |     |        |                 |
| SOURCE     | Dictyostelium discoideum                                  |           |     |        |                 |
| ORGANISM   | Dictyostelium discoideum                                  |           |     |        |                 |
|            | Eukaryota: Mvctozoa: Dictyosteliida: Dictyostelium        |           |     |        |                 |





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:20:52 ; Search time 492.114 Seconds  
(without alignments)  
7597.608 Million cell updates/sec

Title: US-10-070-574-1\_COPY\_574\_1134

Perfect score: 561

Sequence: 1 acgacgcgttttgccttag.....cacagattaagctcagaac 561

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3312346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_21.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004as.\*

13: Geneseqn2004bs.\*

14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 561   | 100.0       | 1146   | 5     | Aaf61989 K. marxia |
| 2          | 87.4  | 15.6        | 2270   | 4     | Abi24848 Drosophil |
| 3          | 87.4  | 15.6        | 5454   | 3     | Aaa70236 Plasmodiu |
| 4          | 87.2  | 15.5        | 4985   | 6     | Abq75107 Anopheles |
| 5          | 87.2  | 15.5        | 4985   | 10    | Acf79720 Mosquito  |
| 6          | 85    | 15.2        | 183    | 12    | Adx61708 Base cont |
| 7          | 85    | 15.2        | 1507   | 13    | Adx50603 Plant ful |
| 8          | 84    | 15.0        | 6642   | 4     | Abi05270 Drosophil |
| 9          | 84    | 15.0        | 10992  | 4     | Abi05242 Drosophil |
| 10         | 84    | 15.0        | 11143  | 4     | Abi12834 Drosophil |
| 11         | 82.2  | 14.7        | 529    | 10    | Abz24320 Sclerotin |
| 12         | 79.4  | 14.2        | 2061   | 3     | Aaa70228 Plasmodiu |
| 13         | 77.6  | 13.8        | 1962   | 14    | Adz04202 Plasmodiu |
| 14         | 77.2  | 13.8        | 357    | 3     | Aaz94801 Soybean m |
| 15         | 77    | 13.7        | 412    | 3     | Abn81085 Shrimp po |
| 16         | 76    | 13.5        | 382    | 3     | Abn81050 Shrimp po |
| 17         | 76    | 13.5        | 8413   | 6     | Abi34497 Human met |
| 18         | 76    | 13.5        | 8413   | 6     | Abi170520 Chemical |
| 19         | 76    | 13.5        | 8413   | 7     | AdS99758 Complemen |

|   |    |      |      |       |    |          |
|---|----|------|------|-------|----|----------|
| C | 20 | 76   | 13.5 | 10640 | 4  | AAD03729 |
| C | 21 | 75.8 | 13.5 | 2892  | 6  | ABN79885 |
| C | 22 | 75.6 | 13.5 | 183   | 6  | ABN79885 |
| C | 23 | 75.6 | 13.5 | 198   | 2  | AAQ71811 |
| C | 24 | 75.6 | 13.5 | 213   | 6  | AAQ71801 |
| C | 25 | 75.6 | 13.5 | 213   | 6  | AAQ71801 |
| C | 26 | 75.6 | 13.5 | 216   | 6  | AAQ71802 |
| C | 27 | 75.6 | 13.5 | 216   | 6  | AAQ71802 |
| C | 28 | 75.2 | 13.4 | 9408  | 3  | AAQ70157 |
| C | 29 | 74.8 | 13.3 | 2886  | 10 | ACC59886 |
| C | 30 | 74.8 | 13.3 | 3300  | 3  | AAZ55699 |
| C | 31 | 74.6 | 13.3 | 567   | 3  | AAA29549 |
| C | 32 | 74.6 | 13.3 | 4225  | 14 | ABE21661 |
| C | 33 | 74   | 13.2 | 18692 | 4  | ABK42795 |
| C | 34 | 74   | 13.2 | 18692 | 9  | ADB60951 |
| C | 35 | 73.8 | 13.2 | 5940  | 3  | AAA70105 |
| C | 36 | 73.4 | 13.1 | 2000  | 10 | ACC61174 |
| C | 37 | 73.4 | 13.1 | 2000  | 10 | ADK63087 |
| C | 38 | 72.4 | 12.9 | 6033  | 3  | AAA70152 |
| C | 39 | 72.2 | 12.9 | 6621  | 3  | AAA70188 |
| C | 40 | 72   | 12.8 | 7143  | 3  | AAA70250 |
| C | 41 | 71.6 | 12.8 | 6198  | 4  | ABL02742 |
| C | 42 | 71.2 | 12.7 | 185   | 3  | AAZ94810 |
| C | 43 | 71.2 | 12.7 | 213   | 3  | AAZ94809 |
| C | 44 | 71.2 | 12.7 | 2364  | 3  | AAA70246 |
| C | 45 | 71.2 | 12.7 | 11922 | 3  | AAA70187 |

## ALIGNMENTS

RESULT 1

AAAF61989

ID AAFA61989 standard; DNA; 1146 BP.

AC AAFA61989;

DT 22-AUG-2001 (first entry)

DE K. marxianus endopolysaccharuronase regulatory region DNA SEQ ID 1.

KW Endopolysaccharuronase; promoter; hepatitis B surface antigen; pectin;

KW polyoma virus VP1; staphylococcal protein A; vaccine; antiviral;

KW antibacterial; regulatory region; ds.

OS Kluyveromyces marxianus.

PN WO200120005-A1.

PD 22-MAR-2001.

PF 05-SEP-2000; 2000WO-EP008662.

PR 10-SEP-1999; 99DE-01043383.

XX (TADP-) TAD PHARMA GMBH.

XX Becher D, Sieketele R, Bartkeviciute D, Sasnauskas K, Doehner L;

XX Salim S;

XX WPI; 2001-244812/25.

XX New promoter from Kluyveromyces marxianus, useful for controlling

XX expression of foreign genes in yeast, e.g. production of vaccine

XX antigens.

XX Claim 1; Page 26; 32pp; German.

XX This invention describes a novel promoter sequence (S1), isolated from

XX Kluyveromyces marxianus. (S1) is a promoter for controlling expression of

XX foreign proteins in yeast, particularly hepatitis B surface antigen,

XX polyoma virus VP1 or staphylococcal protein A for use in vaccines. (S1)

XX is a strong promoter that can be induced by adding pectin to the culture









CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX

QQ Sequence 1507 BP; 393 A; 403 C; 415 G; 296 T; 0 U; 0 Other;  
Query Match 15.2%; Score 85; DB 13; Length 1507;  
Best Local Similarity 82.9%; Pred. No. 1.1e-05;  
Matches 97; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 109 GAAACATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 168  
DB 1363 GAGATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1304  
QY 169 ATTGAAATGTTATTGTTCTTACTGTTCTACTATTATTATTATTATTATTATTATT 225  
DB 1303 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1247

RESULT 8  
ABL05270  
ID ABL05270 standard; cDNA; 6642 BP.  
XX  
AC ABL05270;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10292.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
PI WPI; 2001-556860/75.  
XX  
PI P-PSDB; ABB61167.  
XX  
XX  
PT Genes isolated nucleic acid detection reagent for detecting 1000 or more  
PT proteins from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX  
XX Claim 1; SEQ ID NO 10292; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB101840-AB116175) and the encoded proteins (ABBS57737-  
XX ABB57202). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

[illegible]





```

ABN81085/c
ID ABN81085 standard; DNA; 412 BP.
XX
AC ABN81085;
XX
DT 16-JUL-2002 (first entry)
XX
DE Shrimp polynucleotide SEQ ID NO 89.
XX
KW Giant black tiger prawn; Penaeus monodon; pacific white shrimp;
KW Litopenaeus vannamei; shrimp; microsatellite sequence; genome mapping;
XX Taura Syndrome Virus; TSV; infection; ds.
XX
OS Penaeus monodon.
XX
PN WO200034476-A2.
XX
PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-US029571.
XX
PR 10-DEC-1998; 98US-0111670P.
XX
PA (TUFT ) TUFTS COLLEGE.
XX
PI Alcivar-Warren A, Xu Z, Dhar AK, Fan Y, Meehan D, Garcia DK;
XX WPI; 2000-423422/36.
XX
PT Polynucleotides of shrimp are useful for identifying, mapping and
PT characterizing of the genome of various species of shrimp.
XX
PS Claim 1; Page 98; 120pp; English.
XX
CC The invention relates to an isolated polynucleotide (I) of the giant
CC black tiger prawn, Penaeus monodon or expressed sequence tags of the
CC pacific white shrimp, Litopenaeus vannamei (ABN80997-ABN81172), both
CC containing microsatellites sequences including those P. monodon
CC microsatellite sequences given in GenBank AF077550-AF077598. (I), the
CC complementary sequence or fragment and the encoded polypeptide are useful
CC for mapping of the genome of various species of shrimp. Mapping the
CC genome of Penaeus is useful for determining whether a test shrimp,
CC preferably Litopenaeus vannamei, has a genotype associated with a
CC phenotypic trait such as resistance to Taura Syndrome Virus (TSV)
XX infection
XX
SQ Sequence 412 BP; 213 A; 30 C; 41 G; 128 T; 0 U; 0 Other;
Query Match 13.7%; Score 77; DB 3; Length 412;
Best Local Similarity 76.0%; Pred No. 0.00028;
Matches 95; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
Qy 85 ATATGCAATGTAGAGTAAGTCGTCGAAACCACTATTATTATTATTATTATTATTATT 144
Db 192 ATTATCATTTATCAATAAATACCATCGTTATTATTATTATTATTATTATTATTATT 133
Qy 145 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 204
Db 132 ATTATTGTTATTATTATTCATTTATTATTATTATTATTATTATTATTATTATTATT 73
Qy 205 ATTAT 209
Db 72 ATTAT 68

```

Search completed: February 23, 2006, 01:52:07  
Job time : 494.114 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:40:58 : Search time 3385.83 Seconds  
(without alignments)  
7752.175 Million cell updates/sec

Title: US-10-070-574-1\_COPY\_574\_1134

Perfect score: 561

Sequence: 1 agcagcggtttgtctacag.....cacagattaagctcagaac 561

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 segs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 93.8  | 16.7        | 702    | 10    | CW161301    |
| 2          | 93.4  | 16.6        | 891    | 10    | CW952279    |
| 3          | 92.2  | 16.4        | 579    | 1     | AU038989    |
| 4          | 92.2  | 16.4        | 655    | 4     | AY068594    |
| 5          | 92.2  | 16.4        | 744    | 10    | CW161302    |
| 6          | 92    | 16.4        | 580    | 9     | BZ639429    |
| 7          | 91.8  | 16.4        | 435    | 10    | CW039478    |
| 8          | 91.6  | 16.3        | 832    | 9     | BH391984    |
| 9          | 91.2  | 16.3        | 762    | 10    | CL992030    |
| 10         | 91.2  | 16.3        | 876    | 9     | BZ700706    |
| 11         | 91.2  | 16.3        | 985    | 10    | CG205910    |
| 12         | 90.8  | 16.2        | 701    | 10    | CG200505    |
| 13         | 90.8  | 16.2        | 946    | 10    | CG814583    |
| 14         | 90.6  | 16.1        | 713    | 9     | AZ217550    |
| 15         | 90.6  | 16.1        | 891    | 10    | CW944789    |
| 16         | 90.4  | 16.1        | 637    | 10    | CW293181    |
| 17         | 90.4  | 16.1        | 674    | 10    | CL705363    |
| 18         | 90.4  | 16.1        | 679    | 10    | CW233591    |
| 19         | 90.4  | 16.1        | 843    | 4     | AY068764    |
| 20         | 90.2  | 16.1        | 617    | 3     | BJ394972    |
| 21         | 90    | 16.0        | 693    | 10    | CW189159    |
| 22         | 89.8  | 16.0        | 622    | 10    | CW291169    |

|   |    |      |      |     |    |          |
|---|----|------|------|-----|----|----------|
| C | 23 | 89.6 | 16.0 | 445 | 3  | BJ392338 |
|   | 24 | 89.6 | 16.0 | 474 | 10 | CG106547 |
| C | 25 | 89.6 | 16.0 | 590 | 3  | BJ365627 |
|   | 26 | 89.4 | 15.9 | 480 | 9  | BZ393762 |
| C | 27 | 89.4 | 15.9 | 624 | 10 | CL801804 |
| C | 28 | 89.4 | 15.9 | 711 | 9  | BH969990 |
| C | 29 | 89.2 | 15.9 | 311 | 1  | AU054171 |
| C | 30 | 89.2 | 15.9 | 546 | 9  | BZ999891 |
|   | 31 | 89.2 | 15.9 | 839 | 10 | BX168691 |
|   | 32 | 88.8 | 15.8 | 520 | 8  | DN550639 |
|   | 33 | 88.8 | 15.8 | 541 | 8  | DN550960 |
|   | 34 | 88.8 | 15.8 | 565 | 8  | DN203124 |
|   | 35 | 88.8 | 15.8 | 617 | 8  | DN551066 |
|   | 36 | 88.8 | 15.8 | 673 | 10 | CW024383 |
|   | 37 | 88.8 | 15.8 | 690 | 10 | CW165754 |
|   | 38 | 88.6 | 15.8 | 305 | 1  | AU038805 |
|   | 39 | 88.6 | 15.8 | 750 | 10 | CG033026 |
|   | 40 | 88.6 | 15.8 | 832 | 10 | CG136380 |
|   | 41 | 88.6 | 15.8 | 989 | 11 | CNS032JH |
|   | 42 | 88.4 | 15.8 | 352 | 10 | CW965330 |
|   | 43 | 88.4 | 15.8 | 654 | 10 | CW372547 |
|   | 44 | 88.4 | 15.8 | 687 | 9  | BH401326 |
|   | 45 | 88.4 | 15.8 | 738 | 10 | CL159291 |

## ALIGNMENTS

|            |  |                                |         |                              |   |
|------------|--|--------------------------------|---------|------------------------------|---|
| RESULT 1   | CW161301   | 702 bp.                        | DNA     | linear                       | GSS 29-OCT-2004   |
| LOCUS      | CW161301   | 104_568_11150342_116_36436_061 | Sorghum | methylation filtered library | (LibID: 104) Sorghum bicolor genomic clone 11150342, genomic survey sequence. |
| DEFINITION | CW161301   |                                |         |                              |   |
| ACCESSION  | CW161301   | GI:54853848                    |         |                              |   |
| VERSION    | CW161301   |                                |         |                              |   |
| KEYWORDS   | GSS.   |                                |         |                              |   |
| SOURCE     | Sorghum bicolor (sorghum)  |                                |         |                              |   |
| ORGANISM   | Sorghum bicolor  |                                |         |                              |   |
| REFERENCE  | 1 (bases 1 to 702)   |                                |         |                              |   |
| AUTHORS    | Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holum, H., Roe, B.A., Wiley, G., Korff, I.F., Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddloch, J.A. and Martienssen, R.A.              |                                |         |                              |   |
| TITLE      | Sorghum genome sequencing by methylation filtration  |                                |         |                              |   |
| JOURNAL    | PLOS Biol. 3 (1), e13 (2005)   |                                |         |                              |   |
| PUBLISHED  | 15660154   |                                |         |                              |   |
| COMMENT    | Contact: Bedell JA<br>Orion Genomics, LLC<br>4041 Forest Park Ave, St. Louis, MO 63108, USA<br>Tel: 314 615 6979<br>Fax: 314 615 5975<br>Email: jbedell@oriongenomics.com<br>Plate: 568 row: d column: 14<br>Seq primer: T3 Reverse<br>Class: methylation filtered<br>High quality sequence stop: 702. |                                |         |                              |   |
| FEATURES   | Location/Qualifiers  |                                |         |                              |   |
| source     | 1..702   |                                |         |                              |   |
|            | /organism="Sorghum bicolor"  |                                |         |                              |   |
|            | /mol_type="genomic DNA"  |                                |         |                              |   |
|            | /cultivar="MTx623"   |                                |         |                              |   |
|            | /db_xref="taxon:4558"  |                                |         |                              |   |
|            | /clone="11150342"  |                                |         |                              |   |
|            | /104"  |                                |         |                              |   |
|            | /note="Organ: leaf; Vector: pBCK(-); Site 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5  |                                |         |                              |   |



kb fraction, ligated into HincII-digested pBSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

## ORIGIN

Query Match 16.7%; Score 93.8; DB 10; Length 702;  
Best Local Similarity 70.6%; Pred. No. 2.3e-09;  
Matches 125; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 54 TATTAACTTAGAATTAAGCAACATCTTTGGATGATGATGAGTAAGTCGTTGGAAC 113  
|||||  
Db 512 TATCAAAAAAAGTAATGATACAGCACCAAAATTCAGTTGTTTATTATTATTAT 571  
|||||

QY 114 CATTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTGA 173  
|||||  
Db 572 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT 631  
|||||

QY 174 AATTGTTATTGTTCTTAGTTTCACACTATTATTATTATTCATATTCATGTTATTGACAT 230  
|||||  
Db 632 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAATAT 688  
|||||

RESULT 2  
CW952279 891 bp DNA linear GSS 20-DEC-2004  
LOCUS Tc838\_1\_B02\_T7 Tribolium BAC library Tribolium castaneum genomic,  
DEFINITION genomic survey sequence.

ACCESSION CW952279

VERSION CW952279.1 GI:56735316

KEYWORDS GSS.

SOURCE Tribolium castaneum (red flour beetle)

ORGANISM Tribolium castaneum

REFERENCE 1 (bases 1 to 891)  
AUTHORS Savard J. and Tautz D.  
TITLE Tribolium castaneum BAC-ends sequencing project  
JOURNAL Unpublished (2003)

COMMENT Contact: Savard, J.

Abteilung für Evolutionsgenetik, AG Tautz

Institut für Genetik, Universität zu Köln

Weyertal 121, 50931 Köln, Germany

Tel.: 49 221 470 6911

Fax: 49 221 470 5975

Email: savard@uni-koeln.de

Class: BAC ends.

Location/Qualifiers

1. 891

/organism="Tribolium castaneum"

/mol\_type="genomic DNA"

/strain="GA-2"

/db\_xref="taxon:7070"

/clone\_lib="Tribolium BAC library"

/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; Library constructed by Exelixis Inc."

## FEATURES

source

## ORIGIN

Query Match 16.6%; Score 93.4; DB 10; Length 891;  
Best Local Similarity 66.8%; Pred. No. 2.7e-09;  
Matches 133; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 32 AATGACGAGATTTTGCATAGTATTAACCTTAGAATTAAGCAACATCTTTGGATATGA 91  
|||||

Db 106 AATCAGTTGAATCGTATCTTTGCTTTTATTAATTAATGCAAACTTTATCATGTGAGCT 165  
|||||

QY 92 TGTCAGTAGAGTCGTTTGGAAACCATTTATTATTATTATTATTATTATTATTATTATTA 151  
|||||

Db 166 TTTTCAGTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTA 225  
|||||

QY 152 TTATTATTATTAGTATTATTAAGAAATGTTGTTCTTAGTTTCACTACTATTATTATTC 211  
|||||

Db 226 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTA 285  
|||||

QY 212 ATATTCATGTTATTGACAT 230  
|||||

Db 286 TTATTATTATTATTATTAT 304  
|||||

## RESULT 3

AU038989/c

LOCUS AU038989

DEFINITION Dictyostelium discoideum SS (H.Uruehihara) Dictyostelium

discoideum cDNA clone SSM246, mRNA sequence.

ACCESSION AU038989

VERSION AU038989.1

KEYWORDS EST.

SOURCE Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE 1 (bases 1 to 579)

AUTHORS Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,

Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,

Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

The Dictyostelium developmental cDNA project: generation and

analysis of expressed sequence tags from the first-finger stage of

development

JOURNAL DNA Res. 5 (6), 335-340 (1998)

PUBMED 10048482

COMMENT Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan

Tel: 81-298-53-4664

Fax: 81-298-53-6614

Email: hideko@biol.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'

Location/Qualifiers

1. 579

/organism="Dictyostelium discoideum"

/mol\_type="mRNA"

/strain="AX4"

/db\_xref="taxon:44689"

/clones="SSM246"

/dev\_stage="slug"

/clone\_lib="Dictyostelium discoideum SS (H.Uruehihara)"

## ORIGIN

source

Query Match 16.4%; Score 92.2; DB 1; Length 579;  
Best Local Similarity 70.1%; Pred. No. 4.9e-09;  
Matches 124; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 55 ATTAACCTTAGAATTAAGGCAACATCTTTGGATATGCGATAGAGTAAGTCGTTGGAAC 114  
|||||

Db 334 ATTAATCATTAATGTTTAAATCAATTAATAATTAATTAATTAATTAATTAATTAAT 275  
|||||

QY 115 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 174  
|||||

Db 274 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 215  
|||||

QY 175 ATTGTTATTGTTCTTAGTTTCACTACTATTATTATTATTATTATTATTATTATTATTCAT 231  
|||||

Db 214 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 158  
|||||

## RESULT 4

AY068594/c

LOCUS AY068594

DEFINITION Schmidtea mediterranea clone H.69.1f unknown mRNA sequence.

ACCESSION AY068594

VERSION AY068594.1

KEYWORDS HTC.

SOURCE Schmidtea mediterranea

ORGANISM Schmidtea mediterranea

REFERENCE 1 (bases 1 to 655)

AUTHORS Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Sericata;

Tricladida; Paludicola; Dugesidae; Schmidtea.

```

Class: methylation filtered
High quality sequence stop: 744.
Location/Qualifiers
1. - 744
/organism="Sorghum bicolor"
/mol_type="Genomic DNA"
/cultivar="Atx623"
/db_xref="taxon:4558"
/clone="11150342"
/clone_lib="Sorghum methylation filtered library (LibID:
104)"
/note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBCSK(-) vector
and electroporated into E. coli cells. This is a
methylation filtered library."

```

```

ncp://pianara.med.utah.edu.
Location/Qualifiers
    1..655
        /organism="Schmidtea mediterranea"
        /mol_type="mRNA"
        /strain="CIW4"
        /db_xref="taxon:79327"
        /clone="H.69.lf"

ORIGIN
Query Match          16.4%; Score 92.2; DB 4; Length 655;
Best Local Similarity 74.7%; Pred. No. 4.9e-09;
Matches 115; Conservative 0; Mismatches 39; Indels 0; Gaps 0

QY      77 ATCTTGGATATGCATGTAGTAGTAAGTCGTTCCGAACCATTTATTATTATTATTATTATTA 136
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      322 AATTTTGATGTCNGNTTTATGTAAGCAGGAATAATTATTATTATTATTATTATTATTA 263
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      137 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 196

```

197 CTACTATTATTATTCATATTCATGTTATTGCAT 230  
|| ||||||| ||| |||||  
202 TTATTATTATTATTATTATTATTATTATTAT 169

| RESULT 5   | 744 bp | DNA      | linear        | GSS 29-OCT-2004  |
|------------|--------|----------|---------------|--|
| CW161302/c | 104    | 11150342 | 148 36435 061 | Sorghum methylation filtered library                   |
| LOCUS      | 104    | 11150342 | 148 36435 061 | Sorghum genomic clone 11150342, genomic survey         |
| DEFINITION | 104    | 11150342 | 148 36435 061 | Sorghum bicolor genomic clone 11150342, genomic survey |
| ACCSSION   | 104    | 11150342 | 148 36435 061 | Sorghum bicolor genomic clone 11150342, genomic survey |
| VERSION    | 104    | 11150342 | 148 36435 061 | Sorghum bicolor genomic clone 11150342, genomic survey |
| KEYWORDS   | 104    | 11150342 | 148 36435 061 | Sorghum bicolor genomic clone 11150342, genomic survey |
| SOURCE     | 104    | 11150342 | 148 36435 061 | Sorghum bicolor genomic clone 11150342, genomic survey |
| ORGANISM   | 104    | 11150342 | 148 36435 061 | Sorghum bicolor genomic clone 11150342, genomic survey |
| REFERENCE  | 104    | 11150342 | 148 36435 061 | Sorghum bicolor genomic clone 11150342, genomic survey |

Jones, J.J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,  
 McMenamy, J., Smith, M., Holean, H., Roe, B.A., Wiley, G., Korff, I.F.,  
 Rabinowicz, P. D., Lakey, N., McCombie, W.R., Jeddeloh, J.A. and  
 Martienssen, R.A.  
 Sorghum genome sequencing by methylation filtration  
 PLoS Biol. 3 (1), e13 (2005)  
 15660154  
 Contact: Bedell JA

4441 Forest Park Ave, St. Louis, MO 63108, USA  
Tel: 314 615 6979  
Fax: 314 615 5975  
Email: [jbedello@origenomics.com](mailto:jbedello@origenomics.com)  
Plate: 568 row: d column: 14  
Seq primer: SWfor Forward

## methylation filtered genomic DNA library"

ORIGIN

Query Match 16.4%; Score 92; DB 9; Length 580;  
 Best Local Similarity 72.6%; Pred. No. 5.4e-09;  
 Matches 119; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 67 TTAAGGCAACATCTTTGGATATGTCATGTAGAGTAAGTCGTTGGAACCATTTATTATT 126  
 |||||  
 DB 217 TTGAGAACAAATTAAGTAATGCTAGCAAGGACCATGATATTATTATTATTATT 158  
 |||||

QY 127 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 186  
 |||||  
 DB 157 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 98  
 |||||

QY 187 CTTAGTTCACACTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 230  
 |||||  
 DB 97 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 54  
 |||||

## RESULT 7

CW039478/c  
 LOCUS CW039478 435 bp DNA linear GSS 28-OCT-2004  
 DEFINITION 104 272 1050526 114 30389 Sorghum methylation filtered library  
 (LibID: 104) Sorghum bicolor genomic clone 1050526, genomic survey  
 sequence.

ACCESSION CW039478  
 VERSION 1  
 KEYWORDS GSS.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 435)  
 AUTHORS Bedell J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D.,  
 Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,  
 McMenamy, J., Smith, M., Holsman, H., Roe, B.A., Wiley, G., Korfi, I.F.,  
 Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddlooh, J.A. and  
 Martienssen, R.A.  
 TITLE Sorghum genome sequencing by methylation filtration  
 JOURNAL PLoS Biol. 3 (1), e13 (2005)  
 PUBMED 15660154

COMMENT Contact: Bedell JA  
 Orion Genomics, LLC  
 4041 Forest Park Ave, St. Louis, MO 63108, USA  
 Tel: 314 615 6979  
 Fax: 314 615 5975  
 Email: jbedell@oriongenomics.com  
 Plate: 272 row: f column: 04  
 Seq primer: M13/pUC Forward  
 Class: methylation filtered  
 High quality sequence stop: 435.  
 Location/Qualifiers

FEATURES  
 source  
 1..435  
 /organism="Sorghum bicolor"  
 /mol\_type="genomic DNA"  
 /cultivar="Atx623"  
 /db\_xref="taxon:4558"  
 /clone="1050526"  
 /clone\_lib="Sorghum methylation filtered library (LibID: 104)"  
 /note="Organ: leaf; Vector: pBSK(-); Site 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

## ORIGIN

Query Match 16.4%; Score 91.8; DB 10; Length 435;  
 Best Local Similarity 75.5%; Pred. No. 6e-09;  
 Matches 114; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 80 TTTCGATATGCATGTAGAGTAAGTCGTTGGAACCATTTATTATTATTATTATTATTATT 139  
 |||||  
 DB 183 TATGAACATGTGCTTCGTAATGCTTTTAATACATATTATTATTATTATTATTATTATT 124  
 |||||  
 QY 140 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 199  
 |||||  
 DB 123 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 64  
 |||||  
 QY 200 CTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 230  
 |||||  
 DB 63 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 33  
 |||||

## RESULT 8

BH391984/c  
 LOCUS BH391984 832 bp DNA linear GSS 11-DEC-2001  
 DEFINITION AG-ND-138G17.TR ND-TAM Anopheles gambiae genomic clone  
 AG-ND-138G17, genomic survey sequence.

ACCESSION BH391984  
 VERSION 1  
 KEYWORDS GSS.  
 SOURCE Anopheles gambiae (African malaria mosquito)  
 ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoides;  
 Culicidae; Anophelinae; Anopheles.

REFERENCE 1 (bases 1 to 832)  
 AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J.,  
 Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B.,  
 Gardner, M.J. and Collins, F.H.

TITLE Construction of a BAC library and generation of BAC end  
 sequence-tagged connectors for genome sequencing of the African  
 malaria mosquito Anopheles gambiae

JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)  
 PUBMED 12655398  
 COMMENT Other GSSs: AG-ND-138G17.TF.1  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by  
 F.H. Collins and sequenced by The Institute for Genomic Research  
 (TIGR). The BAC library was generated from A. gambiae PEST strain  
 DNA. All DNA was extracted from newly hatched first instar larvae  
 to minimize the inclusion of DNA from microorganisms that inhabit  
 the gut. The DNA is derived from mixed sexes of larvae. The BAC  
 library was constructed at Texas A&M University BAC Center  
 University, College Station, Texas 77843-2123, USA using a HindIII  
 partial digest.  
 Seq primer: M13 Rev  
 Class: BAC ends.  
 Location/Qualifiers

FEATURES  
 source  
 1..832  
 /organism="Anopheles gambiae"  
 /mol\_type="genomic DNA"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-138G17"  
 /clone\_lib="ND-TAM"  
 /note="Vector: pECBAC1; Site 1: HindIII"

## ORIGIN

Query Match 16.3%; Score 91.6; DB 9; Length 832;  
 Best Local Similarity 71.2%; Pred. No. 6.5e-09;  
 Matches 121; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 61 TTGAATTAAGGCAACATCTTTGGATATGTCATGTAGTAAGTCGTTGGAACCATTTATT 120  
 |||||  
 DB 587 TTATTAATTTATCTTTTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 528  
 |||||







